

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:18:28 ; Search time 4477.84 Seconds
(without alignments)
9887.411 Million cell updates/sec

Title: US-10-627-124-1

Perfect score: 1215

Sequence: 1 atgagccaatcgatcaatt.....tcgtgcgttgacggggca 1215

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hic:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gsa1:*
- 9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.2	4.5	975	8	BZ675691 PUBDH78TD
2	45	3.7	711	5	BW248443 BW248443
3	44.2	3.6	697	2	BE585802 Est#2pt7
4	44	3.6	569	6	CB827487 LjNEST75d
5	44	3.5	1101	9	AL063921 Drosophila
6	41.8	3.4	439	1	AL823229 AL823229
7	41.6	3.4	330	8	AY174775 AY174775
8	40.6	3.3	631	7	CK535040 tswgb0.00
9	40.6	3.3	777	6	CA923083 EST640801
10	40	3.3	747	7	CF347163 AGENCOURT
11	40	3.3	905	7	CN317145 AGENCOURT
12	40	3.3	930	6	CF265727 AGENCOURT
13	39.4	3.2	779	5	BU796025 SJF2DXG07
14	39.2	3.2	585	2	BE432743 EST39272
15	38.8	3.2	705	5	BQ853232 QGB1905.y
16	38.8	3.2	732	5	BU013226 QGJ4B04.y
17	38.6	3.2	643	7	CN086902 EC2BBA27D
18	38.6	3.2	662	7	CF388973 RTDR2.16
19	38.6	3.2	670	7	CF477876 RTW3.13
20	38.6	3.2	682	7	CO369247 RTK1.45H
21	38.6	3.2	688	7	CN086901 EC2BBA27D
22	38.6	3.2	689	7	CF479071 RTW3.21
23	38.6	3.2	712	6	CF113752 Shultzomi
24	38.6	3.2	714	7	CR420363 CR420363

ALIGNMENTS

RESULT 1
BZ675691
LOCUS PUBDH78TD ZM 0.6_1.0_KB Zea mays genomic clone ZMMBTA029M12,
DEFINITION genomic survey sequence.
ACCESSION BZ675691
VERSION BZ675691.1 GI:28226790
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 975)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J., and
Bennetzen,J.
Maize Genomics Consortium
TITLE Unpublished (2003)
JOURNAL
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
source
1. .975
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA029M12"
/clone_lib="ZM 0.6_1.0_KB"
/note="Vector: pCR4-TOPO; Site:1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

Query Match 4.5%; Score 54.2; DB 8; Length 975;
Best Local Similarity 46.2%; Pred. No. 5.4e-05;
Matches 222; Conservative 0; Mismatches 253; Indels 6; Gaps 1;

QY 514 CCTGAAGCGCGGCACATTCGTTATGTCACGACAGCTTCCCAATACTGTGCGG 573
|||||
DB 125 CCTGAAGCGCGGCATCAGACTCAGGCTCATCTGCGGAGCTGCGGCATACATCAAAAAC 184
|||||

CO172749 ND1.31.D
CF394958 RTDS2.8.H
CR420364 CR420364
CF390685 RTDR2.20
CO593228 AGENCOURT
CG970675 MBEKLS1TR
CG968903 MBEHWA0TF
BH614276 1C22/2B8
BH614466 2B8.Cont1
BG702387 602583676
BO866560 QGC8F03.Y
B1194654 602946381
CL625541 OR_BBA002
AQ690680 nbxb0082B
CL839054 OR_CBA007
CL736355 OR_BBA007
AQ259669 nbxb0023D
CO107177 GR_EB003
CO103973 GR_EB003
CO072687 GR_Ea31P
CO082104 GR_Ea461

3.2 720 7 CO172749
3.2 726 7 CF394958
3.2 732 7 CR420364
3.2 740 7 CF390685
3.2 771 7 CO593228
3.2 902 9 CG970675
3.2 903 9 CG968903
3.1 663 8 BH614276
3.1 663 8 BH614466
3.1 790 4 BG702387
3.1 728 5 BO866560
3.1 737 4 B1194654
3.1 505 9 CL625541
3.1 768 8 AQ690680
3.1 775 9 CL839054
3.1 789 9 CL736355
3.1 801 8 AQ259669
3.1 805 7 CO107177
3.1 839 7 CO103973
3.1 856 7 CO072687
3.1 922 7 CO082104

25 38.6
26 38.6
27 38.6
28 38.6
29 38.6
30 38.4
31 38.4
32 38.2
33 38.2
34 38.2
35 38
36 37.8
37 37.6
38 37.6
39 37.6
40 37.6
41 37.6
42 37.6
43 37.6
44 37.6
45 37.6

QY 574 GTCAATTAAGAACGTCGCGCTCAATCCGGATCAGATTTAAATTTCCGATCCTATGTACTTCT 633
DB 185 CTGGTGGCGAAGCTCGACCGAACCAACGCGATGACTTCATGACCGCCTGACCCAAAGCC 244
QY 634 GAATATGAAGGATGGCGTGTGCGACAAAGATATATCTGCACCTGATTTCTTAATGTGCTG 693
DB 245 GACGAAAAAGGCGAGTCCCTCAACGAGGAGCAATTGCTGCTGTGTGCAATTCGATCCTG 304
QY 694 TTAGCGCAACCGAACCGGCTGATAGACGCTGGCAGTATCATTCACATTTGCTCAAC 753
DB 305 GTCGCGGTTAGCAAGGCGAGTCTGCGCAGATCCCAACTTATCAGGTGCTGTGGAC 364
QY 754 AATCTGTAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCCGAGAGCCATTGGC 813
DB 365 AACCCGCGCATGGCAACAGCTCAAGGCCGATCCGAGCAGATACCGGAGCGGCTCGNA 424
QY 814 GAGACATTGCGTTATAAACCCG-----CGGTTACGCTGATTTCCGCGGCGAGCTGTCCCA 867
DB 425 GAGTGTCTGCGCTACATCCCGCTGGCCTCGCGGCGATGTTTGTGCACTACGCGCTGGAA 484
QY 868 GATACAGTGGTGGCGGTATGAAATCAAAAAAGATACGATTTGTTTTTGTATGATCGGT 927
DB 485 GACATTTCAAGTTCGGTGAACCCCTGGTGGCCGAGGCGACGCGTGTTCGCCCTCGATCGGC 544
QY 928 GCGGCTAACCGGAGCCCTGAAGCATTTGAACAGCCTGACGCTGTTAAATATTCATCGGAA 987
DB 545 GCGGCNACCATGACCGCGCCGCTTTGAAAACCCACAGCGCCTGGACCTGCAACGGAT 604
QY 988 G 988
DB 605 G 605

RESULT 2

BW248443
LOCUS BW248443 711 bp mRNA linear EST 09-NOV-2002
DEFINITION BW248443 Nori Satoh unpublished cDNA library, tailbud embryo Ciona intestinalis cDNA clone citb07m23 5', mRNA sequence.
ACCESSION BW248443
VERSION BW248443.1 GI:24828361
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Ciona.
REFERENCE 1 (bases 1 to 711)
AUTHORS Satou,Y., Shin-i.T., Kohara,Y. and Satoh,N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satohe@acidian.zool.kyoto-u.ac.jp.

FEATURES

source
1. .711
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="citb07m23"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_lib="Nori Satoh unpublished cDNA library, tailbud embryo"

ORIGIN

Query Match 3.7%; Score 45; DB 5; Length 711;
Best Local Similarity 47.3%; Pred. NO. 0.037;
Matches 169; Conservative 0; Mismatches 185; Indels 3; Gaps 1;
QY 773 ATGTTTGGCTGACCGCTTCGTTAGTTCCGAGAGCCATTGCGGAGACATTCGTTATTAAC 832

DB 206 ATGATCTGTCCAACCTCCACCTTACATTATGATCAAAAGAAAGTTTGGCCCAATATC 265
QY 833 CGCGGTTTCAGTATTCGCGGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAA 892
DB 266 CACCAGTTTCCAATCATTTTTCGTTAAACTTAAACAAAGATATTGAAGTTGATGGAAAGACCA 325
QY 893 TCAAAAAGATACGATGTTTTCGTTATGATCGGTGCGCTAACCGGGACCTCGAAGCAT 952
DB 326 TTGTGAAGACACCAATGTTGTTCTACATATTATGATTATCATCATCAGGAGAGTTTT 385
QY 953 TTGAACAGCCTGACGTTTAAATATTCATCGGAAGATCTTGGTATCAAGAGCGCTTTTA 1012
DB 386 GGAAGGATCCTCATATATTGATCCAAGTCGTTCCAGCAGGAAAAACATGAAAAACATCA 445
QY 1013 GCGGCGCGCGCGGATCTCGCTTCGGATCCGGATTCATTAACCTGTGTAGGAGAGCTT 1072
DB 446 ATAGTTATGCTT---ATGTACCTTTCTCTGCTGGCCCAAGAAACTGCATCGGCGCAAAAAT 502
QY 1073 TTGCCAAAAGCAATCGAAATGTTAGCTAATATTGTGCTGATAGATGCGGAATA 1129
DB 503 TTGCATGAACAAGATGAAATTTGCGTTGCTCAGGTGTTGAGGCAATTCAGATTA 559

RESULT 3

BW585802/c
LOCUS BW585802 697 bp mRNA linear EST 17-AUG-2000
DEFINITION BW585802 Nori Satoh unpublished cDNA library, tailbud embryo Ciona library Triticum aestivum cDNA clone Est#2pT7_B07_b7_057, mRNA sequence.
ACCESSION BW585802
VERSION BW585802.1 GI:9838834
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 697)
AUTHORS Fellers,J.P., Li,W.L., Hill-Ambroz,K., Matthews,A. and Gill,B.S.
TITLE The structure and function of the expressed portion of the wheat genomes - Kansas State University. Fusarium graminearum infected spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: John Fellers
US Department of Agriculture, Agriculture Research Service, Plant Science and Entomology Unit
Dept. of Plant Pathology, 4006 Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA
Tel: 785-532-2367
Fax: 785-532-6167
Email: jpf@alfalfa.ksu.edu
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: 17.

FEATURES

source
1. .697
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Sumai3"
/db_xref="taxon:4565"
/clone="Est#2pT7_B07_b7_057"
/tissue_type="Spike"
/dev_stage="Adult plant"
/lab_host="E. coli JM109"
/clone_lib="KSU wheat Fusarium graminearum infected spike cDNA library"
/note="Vector: pGEM-T easy; Site_1: SacII; Site_2: SpeI;
Plants were grown in the greenhouse. Spikes were sprayed with Fusarium graminearum (at what stage). Total RNA, and poly(A) RNA were prepared from infected spikes. cDNA was prepared using the SmartTM PCR cDNA synthesis kit from Clontech. cDNA was cloned into the pGEM-T easy vector

```

ORIGIN
Query Match 3.6%; Score 44.2; DB 2; Length 697;
Best Local Similarity 51.2%; Pred. No. 0.066;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 775 GTTTGGCTGACGCTGTTAGTTCGAGAGCAATTCGCGAGACATGCGGTATATAACCG 834
Db 617 GATCTTGGCGAACTCCACTACTTGGCGAGTGTCAAGGAGGTTCTTAGATTGCATCCA 558

QY 835 CCGGTTCACTGATCCGGCGCAGCTGTCACAGATACAGTGTCCGGCGGTATGGAATC 894
Db 557 CTGGTCCGCTTAATCCCTCGAGACTAGAGAGATTGCAATATATGAGGTATGACATT 498

QY 895 AAAAAAGATACGATGTTTTTTTGTATGATCGGTGCGGCTAAACCGGACCTCGAAGCATTT 954
Db 497 CCNAAAGATACAAATATATTCATTATACCTTTGCAACATCCCGGATCTCAATATTGG 438

QY 955 GAACAGCCTGAGCTGTTTAAT 975
Db 437 GACAGTCTCGAAGTGTTTAAT 417

RESULT 4
CB827487
LOCUS
DEFINITION
LjNEST75d12r Lotus japonicus nodule library 5 and 7 week-old Lotus
corniculatus var. japonicus cDNA 5', mRNA sequence.
ACCESSION
CB827487
VERSION
CB827487.1 GI:29966503
SOURCE
Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;
Lotus.
REFERENCE
1 (bases 1 to 569)
Colebatch,G., Freund,S., Trevisan,S. and Udvardi,M.
Lotus japonicus root nodule ESTs: tools for functional genomics
Unpublished (2000).
CONTACT: Udvardi MK
Molecular Plant Nutrition
Max Planck Institute of Molecular Plant Physiology
Am Muehlenberg 1, 14476 Golm, Germany
Fax: 49 331 567 8250
Email: udvardi@pimp-golm.mpg.de
Seq primer: T7
High quality sequence stop: 569.
FEATURES
Location/Qualifiers
1..569
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/cultivar="Gifu (B-129)"
/db_xref="taxon:34305"
/dev_stage="5 and 7 week-old plants"
/clone_lib="Lotus japonicus nodule library 5 and 7
week-old"
/note="Organ: Nodule; Vector: pSPORT1; Site 1: Sali;
Site 2: NotI; The library was prepared using mRNA
extracted from nodules of 5 and 7 week-old Lotus plants.
Nodules were induced by, and contained Mesorhizobium
strain R7A."
ORIGIN
Query Match 3.6%; Score 44; DB 6; Length 569;
Best Local Similarity 52.8%; Pred. No. 0.071;
Matches 95; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 804 AGCATTCGGAGACATTCGGTTATTAACCGCGGTTCAAGCTGATTCGCGGCGGACGTGTC 863
Db 389 AGTTATAAAGAGACACTCAGATTTTATCGCTGGCGCACTAATACCAAGAGAAACAAT 448

from Promega."
QY 864 CCAAGATACAGTGGTCCGGCGGTATGAAATCAAAAAGATACGATTTTGTGTATGAT 923
Db 449 TAAAAATATTATCGTAGATGGGTATGAAATACCAAGCAACAATGTTTATGTAACGT 508

QY 924 CGTCCGGCTACCGGACCCCTGAAGCATTTTGAACAGCCCTGACGTTGTTTAATATTCATCG 983
Db 509 TTGGGCGCATCCATAGGATCCCTGAGGCTTGGAAAGACCCACACAGGTTTAATCCTGATAG 568

RESULT 5
CNS00396/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL063921
VERSION
AL063921.1 GI:4941778
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Ooegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain v2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN
Query Match 3.5%; Score 42; DB 9; Length 1101;
Best Local Similarity 14.6%; Pred. No. 0.37;
Matches 79; Conservative 240; Mismatches 222; Indels 1; Gaps 1;

QY 244 GAAAAGAACATCTCCCAAGAGAGATTTAGTCAGAGCTTTATCGTGACGCACTG 303
Db 1101 GKARRWGDDTDRDKDDMTKWWTKWTKRADDRRWAGDADRWAMDGAGTWTATW 1042

QY 304 GATCATCTCTCCATTGATTAACAAAATCGAGAAACTGTGTACGCTTATCTTGA 363
Db 1041 WWWWWATWDTWWDKWWMTAAKTDATWMTWTRADWRADRGKRDRAATDAG 982

QY 364 AGAGGGAAGATGATCTCGTCAATGATTTTGAAGACGCTTTCGCGTGCACGATG 423
Db 981 AGRDGRKXKDKDKDGDGDDKGGKKKAAKAAKWTAKWDDWDDWDDGAKDKR 922

QY 424 GATCATCTCGGCTGGATATAAGAGACCATGAAAAAATCTCTGAGTGGCAGCGGAGTT 483
Db 424 GATCATCTCGGCTGGATATAAGAGACCATGAAAAAATCTCTGAGTGGCAGCGGAGTT 483
```


LOCUS CK535040 631 bp mRNA linear EST 15-JAN-2004
DEFINITION rswd00_003392.y1 swg Bombyx mori cDNA, mRNA sequence.
ACCESSION CK535040
VERSION CK535040.1 GI:40919494
KEYWORDS Bombyx mori (domestic silkworm)
SOURCE Bombyx mori
ORGANISM Bombyx mori

REFERENCE 1 (bases 1 to 631)
AUTHORS Xia,Q., Zhou,Z., Lu,C., Cheng,D., Dai,F., Li,B., Zhao,P., Zha,X.,
Cheng,T., Chai,C., Pan,G., Xu,J., Liu,C., Lin,Y., Qian,J., Hou,Y.,
Wu,Z., Li,G., Pan,M., Li,C., Shen,Y., Lan,X., Yuan,L., Yang,G.,
Wan,Y., Zhu,Y., Yu,M., Shen,W., Wu,D., Xiang,Z., Yu,J., Wang,J.,
Li,R., Li,H., Li,G., Su,J., Wang,X., Li,G., Zhang,Z., Li,J.,
Zhang,Q., Wei,N., Xu,J., Sun,H., Dong,L., Liu,D., Zhao,S., Zhao,X.,
Meng,Q., Lan,F., Fang,L., Li,Y., Wang,J., Ye,J., Wong,G.K.-S.,
Yang,H., Shi,J., Wu,Q., Li,C., Li,D., Sun,Y., Zhang,Z., Yang,Z.,
Huang,X., Xi,Y., Qi,Q., He,D., Liang,H., Zhang,X., Wang,Z., Li,W.
and Cao,Y.

TITLE A draft sequence for the genome of the domesticated silkworm
(Bombyx mori)

JOURNAL Unpublished (2004)
COMMENT Contact: Yang Huan Ming
Beijing Genomics Institute
Chinese Academy of Sciences
Beijing Airport Industrial Zone B6, Beijing 101300, P.R.China
Email: hyang@igtp.ac.cn.

FEATURES
Location/Qualifiers
1..831
/organism="Bombyx mori"
/mol_type="mRNA"
/strain="DaZhao (P50)"
/db_xref="taxon:7091"
/sex="male"
/tissue_type="testis"
/dev_stage="5th-instar day-3 larva"
/clone_lib="swg"
/note="Vector: pBluescript II SK(+)"

ORIGIN
Query Match 3.3%; Score 40.6; DB 7; Length 631;
Best Local Similarity 48.9%; Pred. No. 0.85;
Matches 109; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 768 GAATGATGTTTGGCTGACCGTTCTGTTACTTCGAGAGCCATTGCGGACACATTGCGTTA 827
Db 189 GTACAAATGAACGGGACAAATGAAGTTCTGAGTGGGTCAATCAAGGAATCGTTGCGGAT 248

QY 828 TAAACCGCCGGTTTCAGCTATTCGCGGCGAGCTGTCCCAAGATACAGTGGTGGCGGAT 887
Db 249 GTATCCATCGGTACCGCTGATCGACGTTGGATCAAAAGATGCTGAGGTGGAGTCT 308

QY 888 GGAATCAAAAAGATACGATTTGTTTGTATGATCGTGGCGGTAAACCGGACCCCTGA 947
Db 309 GAAACTATCTAAAGGAACATCGTGGGTGGTTAAACATATTCCAGATGCATCGAAATCCAGA 368

QY 948 AGCATTTGAACGCTGAGCTGTTTATATTATTCGCGGAAGAT 990
Db 369 AGGGTTGAGNAGCCCTCGAATTCTTCTCTGAACGATTCGAT 411

RESULT 9
CA923083/c
LOCUS CA923083 MTUS Medicago truncatula cDNA clone MTUS-62F12, mRNA
DEFINITION EST640801
VERSION CA923083
ACCESSION CA923083.1 GI:27410013
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 777)
AUTHORS VandenBosch,K., Endre,G., Silverstein,K., Town,C.D., Van Aken,S.,
Utterback,T., Cheung,F. and Fraser,C.M.
TITLE The Medicago truncatula 6K unigene set: cDNA clones selected and
re-arrayed from various libraries
JOURNAL Unpublished (2002)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
Alias Clone NF013H10IN
TIGR sequence name: MTUCJ72TV
More information is available at: www.medicago.org
Seq primer: (gtA AtA CGA Ctc Act AtA ggg C).

FEATURES
Location/Qualifiers
1..777
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone_lib="MTUS-62F12"
/tissue_type="mixed tissues"
/dev_stage="various stages"
/lab_host="XLOLR"
/clone_lib="MTUS"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XLOLR cells."

ORIGIN
Query Match 3.3%; Score 40.6; DB 6; Length 777;
Best Local Similarity 52.7%; Pred. No. 0.91;
Matches 88; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 807 CATTGCGGAGACATTCGCGTTATAAACCGCGCTTCAGCTGATTCGCGGCGAGTGTCCCA 866
Db 471 CATAAAGAACACATGAGGCTTCATCCCGATACCGATGATTATGAGGAAGGATGGA 412

QY 867 AGATACAGTGTGTCGCGGTATGGAATCAAAAAGATACGATTTGTTTGTATGATCGG 926
Db 411 AGATTGTGTGGTTGATGGTTAAACATGATTCGAAAGGCTCAATGGTTTGTGTAACATTG 352

QY 927 TCGCGTAAACCGGACCCCTGAAGCATTTGAAACAGCTGACGTGTTTA 973
Db 351 GGCTATGGCAAGGAGCCCAAGATATGGGAAACCCATTAGAGTTTA 305

RESULT 10
CF347163
LOCUS CF347163 747 bp mRNA linear EST 18-AUG-2003
DEFINITION AGENCOURT 15223896 NIH_ZGC_10 Danio rerio cDNA clone IMAGE:7002517
5', mRNA sequence.
ACCESSION CF347163
VERSION CF347163.1 GI:33789768
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.


```

source
1. .930
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:6997518"
/tissue_type="whole body"
/lab_host="DH10B"
/clone_lib="NIH ZGC 7"
/notes="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV; Bulk tissue was collected from a whole adult individual from the Tuebingen strain. 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was cloned into the Not I and EcoRV sites of pExpress-1. Library was size-selected for >1 kb fragments and normalized. A non-normalized version of this library is also available (NIH ZGC 10). Library was constructed by Open Biosystems (Huntsville, AL)"

ORIGIN
Query Match 3.3%; Score 40; DB 6; Length 930;
Best Local Similarity 58.3%; Pred. No. 1.5;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 547 GAACAGCTTCCCATATGATCGCGTCATTAAAGACGTCGCTCAATCCGGATCA 606
DB 345 GAACCTCATTAACCAAACTGTGTGCGGCCCTGATTAAATGTTTGACGGCGCGCATTTA 404
QY 607 GATTTAATTCGATCCTATGCTACTTCTGTAATATGAAGGATCGCGTGTGCGACAAAGGAT 666
DB 405 GCCTAGTAGTGCACATGTCGTAGATATTAAGTCATGTCGATGCGAAAGAGAT 464

RESULT 13
LOCUS BU796025 779 bp mRNA linear EST 23-OCT-2003
DEFINITION SJF2DXG07 SJF Schistosoma japonicum cDNA similar to pIRJCL196 protein phosphatase inhibitor 2 - rabbit, mRNA sequence.
ACCESSION BU796025
VERSION BU796025.1 GI:28353032
KEYWORDS EST.
SOURCE Schistosoma japonicum
ORGANISM Schistosoma japonicum
REFERENCE Hu, W., Yan, Q., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R., Wang, Z. J., Rong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J., Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J., McManus, D. P., Xue, C. L., Feng, Z., Chen, Z. and Han, Z. G. Evolutionary and biomedical implications of a Schistosoma japonicum complementary DNA resource Nat. Genet. 35 (2), 139-147 (2003)
JOURNAL 22879925
MEDLINE 12973349
PUBMED
COMMENT Contact: Zeguaguan Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919 (ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn.

FEATURES
Location/Qualifiers
1. .779
/organism="Schistosoma japonicum"
/mol_type="mRNA"
/db_xref="taxon:6182"
/sex="female"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="rabbits"
/clone_lib="SJF"

source
Query Match 3.2%; Score 39.4; DB 5; Length 779;
Best Local Similarity 62.9%; Pred. No. 2.2;
Matches 61; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1064 GAGCAGCTTTTGCACAAACGAAATCGAAATGTAGCTAATATTGTGCTGGATAAGATGC 1123
DB 348 GAGAAATTCGCGAAACGAAACGAAACATTTATGATGATTTTGGCTGTAAACTGCG 407
QY 1124 GGAATATCAGATTAGAGGAAGATTTTGTATTGCTGA 1160
DB 408 TAAAGACAGTTTACAAATATGATGATGATGATGCGGA 444

RESULT 14
LOCUS BE432743 585 bp mRNA linear EST 18-MAY-2001
DEFINITION EST399272 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA clone cLEGI019, mRNA sequence.
ACCESSION BE432743
VERSION BE432743.1 GI:9430586
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 585)
AUTHORS Alcala, J., Vrebalov, J., White, R., van der Hoeven, R. S., Holt, I. E., Liang, F., Hansen, T. S., Craven, M. B., Bowman, C. L., Ronning, C. M., Nierman, W., Fraser, C. M., Martin, G. B., Giovannoni, J. J. and Tanksley, S. D. Generation of ESTs from tomato fruit tissue, breaker stage Unpublished (2000)
JOURNAL Contact: CUGI
COMMENT Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html 5 prime sequence.

FEATURES
Location/Qualifiers
1. 585
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEGI019"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit, TIGR"
/notes="Vector: pBluescriptSMCquad; Site 1: EcoRI; Site 2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopen accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

ORIGIN
Query Match 3.2%; Score 39.2; DB 2; Length 585;
Best Local Similarity 49.1%; Pred. No. 2.3;
Matches 104; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 804 AGCCATTGCGGAGACATTGCGTTATAAACCGCGGTTCAGCTGATTCGCGGAGCTGTC 863
DB 10 AGTTATAAAGAGACATTGAAATTTGATCCCGCTCCATTATAACCAAGAGACAAT 69
QY 864 CCAAGATACAGTGTGCGGGTATGGAATCAAAAAGATAGATGTTTTTGTATGAT 923
DB 70 GGAATAATGCATCATAGATGATGATGAATCAACAGAAATCTCCAGTTTCTATCAACGC 129
QY 924 CGGTGGGCTAACCGGACCCCTGAAGCATTGAAACAGCCTGACGTGTTAAATTCATCG 983
DB 130 GTGGGCAATAGGTAGAGATCCTGAATGTTGGGAAATCCAGAGAGATTTTCTCTGAGAG 189

```

OY 984 GGAAGATCTTGGTATCAAGAGCGCTTTAGCG 1015
| | | | | | | | | | | | | | | | | | | | | |
Db 190 ATTGTGATGTGAACACAAATGGTATTGTCG 221
| | | | | | | | | | | | | | | | | | | | | |

Search completed: December 10, 2004, 18:13:17
Job time : 4485.84 secs

RESULT 15
BQ853232
LOCUS BQ853232
DEFINITION QGB1905.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
ACCESSION QGB1905, mRNA sequence.
VERSION BQ853232
KEYWORDS BQ853232.1 GI:22238697
SOURCE EST.
ORGANISM Lactuca sativa
Lactuca sativa

REFERENCE 1 (bases 1 to 705)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
Lai,Z., Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
JOURNAL http://compgenomics.ucdavis.edu/
COMMENT Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig2847, see http://cgpdb.ucdavis.edu/
for details.

Plate: QGB1 row: 9 column: 05.
FEATURES Location/Qualifiers
1..705
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QGB1905"
/lab_host="E.coli"
/clone_lib="QG_ABCDI lettuce salinas"
/note="Vector: pBRCDNASFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_TISSUE=chemical induction
TAG_LIB=QG_ABCDI lettuce salinas
TAG_SEQ=TGTAGCCGGG"

ORIGIN
Query Match 3.2%; Score 38.8; DB 5; Length 705;
Best Local Similarity 52.5%; Pred. No. 3.2;
Matches 85; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
OY 823 CGTTATAAACCCGGGTTCCAGTGTATTCGGCGCAGCTGCCAAGATACAGTGGTCGGC 882
| | | | | | | | | | | | | | | | | | | | | |
Db 508 CTTTCATCCCCCGCTCCACCCCTAATTCAGAGCCTCCAGCCAGATACAAAGTCATG 567
| | | | | | | | | | | | | | | | | | | | | |
OY 883 GGTATGGAATCAAAAAAGATACGATTGTTTTTTTGTATGATCGGTGGGCTAACCGGGAC 942
| | | | | | | | | | | | | | | | | | | | | |
Db 568 GGCTACGACATCAAGAAAGAACTAGGGTTATCATTAATGATGGGCTATTCAAGAGAC 627
| | | | | | | | | | | | | | | | | | | | | |
OY 943 CCTGAAGCATTTGAACAGCGCTGACGGTGTATTAATATTCATCGG 984
| | | | | | | | | | | | | | | | | | | | | |

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	67.8	5.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 2	67.8	5.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 3	53.6	4.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 4	53.6	4.4	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 5	50	4.1	47981	4	US-09-679-279-1	Sequence 1, Appl
C 6	49.6	4.1	6085	3	US-09-029-603-4	Sequence 4, Appl
C 7	47.6	3.9	1209	4	US-09-724-797-35	Sequence 35, Appl
C 8	44.6	3.7	8478	3	US-08-676-818-1	Sequence 1, Appl
C 9	44.6	3.7	8478	3	US-09-407-549-1	Sequence 1, Appl
C 10	43	3.5	1209	3	US-09-105-537-21	Sequence 21, Appl
C 11	43	3.5	13613	3	US-09-105-537-3	Sequence 3, Appl
C 12	43	3.5	38506	3	US-09-320-878-19	Sequence 19, Appl
C 13	43	3.5	38506	4	US-09-141-908-1	Sequence 1, Appl
C 14	43	3.5	38506	4	US-09-657-440-19	Sequence 19, Appl
C 15	42.4	3.5	37556	2	US-08-576-626A-1	Sequence 1, Appl
C 16	41	3.4	37030	4	US-08-311-731A-25	Sequence 25, Appl
C 17	40.6	3.3	1251	3	US-09-105-537-38	Sequence 38, Appl
C 18	40.6	3.3	5970	3	US-09-320-878-21	Sequence 21, Appl
C 19	40.6	3.3	5970	4	US-09-141-908-11	Sequence 11, Appl
C 20	40.6	3.3	5970	4	US-09-657-440-21	Sequence 21, Appl
C 21	37.8	3.1	1212	6	5212296-17	Patent No. 5212296
C 22	37.4	3.1	1735	1	US-08-102-863-10	Sequence 10, Appl
C 23	37.4	3.1	1735	5	PCR-US92-10885-10	Sequence 10, Appl
C 24	37	3.0	77536	4	US-09-410-551B-1	Sequence 1, Appl
C 25	37	3.0	77536	4	US-09-540-316B-1	Sequence 1, Appl
C 26	36	3.0	536165	2	US-09-214-808-1	Sequence 1, Appl
C 27	35	2.9	4084	2	US-08-568-459A-1	Sequence 1, Appl

QY 948 AGCATTTGAACAGCTGACGTTTAAATATTCATCGGAA 987
Db 874566 CGTGTTCGACGGCGGAGTTCGATATCACCGGAAA 874605

RESULT 4

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 4.4%; Score 53.6; DB 3; Length 4411529;
Best Local Similarity 47.4%; Pred. No. 7.3e-05;
Matches 161; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY 648 GCGCGTGTGCGACAGGATATCTCGCACTGATTTCTTAATGTCTGTAGCGCGAACGGA 707
Db 872136 GCGCGTGTGCGACGCGAAGTACTCTTCTTCAATCTACTGTTGAGCGCGCGCGGA 872195
QY 708 ACCGGCTGATAGACGCTGCGACTGATGATACCAATTTGCTCAACAATCCTGAGCAGAT 767
Db 872196 AACCAACCGTAACTCCATNTCCGCGGGCTGCTGGCGTGGCGAGAACCTTGACCACT 872255
QY 768 GAATGATGTTGGCTGACGCTGCTAGTTCGAGAGCATTGCGGACATTCGCGTTA 827
Db 872256 GCAACGCTGCGAAGCGATTTTGGTGTGCGGCTGCGATCGAAGATCGTGAGGTG 872315
QY 828 TAAACGCGGTTTACGCTGATCCGCGGAGCTGCTCCCAAGATACAGTGTGCGCGGTAT 887
Db 872316 GAGTCGCGCTCACCATCGAAGCGGCGACGGGTCCCGTCCGTCAGCTGCGCGCGCA 872375
QY 888 GGAAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGGCTAAACCGGACCCCTGA 947
Db 872376 GCGATCGAGCGGGTCAGAGGTTGTTGTTGGAGGGCTCGGCCAACCGTGATCCAG 872435
QY 948 AGCATTTGAACAGCTGACGTTTAAATATTCATCGGAA 987
Db 872436 CGTGTTCGACCGCGGAGTTCGATATCACCGGAAA 872475

RESULT 5

US-09-679-279-1/c
; Sequence 1, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; APPLICANT: Volchegursky, Yanina
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
; TITLE OF INVENTION: Genes and Uses Thereof
; FILE REFERENCE: 300622004700
; CURRENT APPLICATION NUMBER: US/09/679,279
; CURRENT FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/158,305
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 60/190,024

; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 47981
; TYPE: DNA
; ORGANISM: Micromonospora megalomicea
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(144)
; OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
; OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (928)...(2061)
; OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (2072)...(3382)
; OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
; OTHER INFORMATION: TDP-megosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (3462)...(4634)
; OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase,
; OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (4651)...(5775)
; OTHER INFORMATION: megDII, deoxysugar transaminase (eryCI, DnrJ homolog),
; OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
; OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (5822)...(6595)
; OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
; OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (6592)...(7197)
; OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmU hon
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
; OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (7220)...(8206)
; OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmV homolog),
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (8228)...(9220)
; OTHER INFORMATION: megBII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
; OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (9226)...(10479)
; OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (10483)...(11424)
; OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
; OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (12181)...(22821)
; OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (12505)...(13470)
; OTHER INFORMATION: megAI, AT-L
; NAME/KEY: misc feature
; LOCATION: (13576)...(13791)
; OTHER INFORMATION: megAI, ACP-L
; NAME/KEY: misc feature
; LOCATION: (13849)...(15126)
; OTHER INFORMATION: megAI, KSI
; NAME/KEY: misc feature
; LOCATION: (15427)...(16476)
; OTHER INFORMATION: megAI, ATI

; NAME/KEY: misc feature
; LOCATION: (1715)...(17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc feature
; LOCATION: (17947)...(18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc feature
; LOCATION: (18258)...(19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc feature
; LOCATION: (19876)...(20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc feature
; LOCATION: (21517)...(22053)
; OTHER INFORMATION: megAI, KR2
; NAME/KEY: misc feature
; LOCATION: (22318)...(22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867)...(33555)
; OTHER INFORMATION: megAI, SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (22957)...(24237)
; OTHER INFORMATION: megAI, KS3
; NAME/KEY: misc feature
; LOCATION: (24544)...(25581)
; OTHER INFORMATION: megAI, AT3
; NAME/KEY: misc feature
; LOCATION: (26230)...(26733)
; OTHER INFORMATION: megAI, KR3 (inactive)
; NAME/KEY: misc feature
; LOCATION: (26958)...(27258)
; OTHER INFORMATION: megAI, ACP3
; NAME/KEY: misc feature
; LOCATION: (27393)...(28590)
; OTHER INFORMATION: megAI, KS4
; NAME/KEY: misc feature
; LOCATION: (28897)...(29931)
; OTHER INFORMATION: megAI, AT4
; NAME/KEY: misc feature
; LOCATION: (29953)...(30477)
; OTHER INFORMATION: megAI, DH4
; NAME/KEY: misc feature
; LOCATION: (31396)...(32244)
; OTHER INFORMATION: megAI, ER4
; NAME/KEY: misc feature
; LOCATION: (32257)...(32799)
; OTHER INFORMATION: megAI, KR4
; NAME/KEY: misc feature
; LOCATION: (33052)...(33312)
; OTHER INFORMATION: megAI, ACP4
; NAME/KEY: CDS
; LOCATION: (33666)...(43271)
; OTHER INFORMATION: megAI, SEQ ID NO: 15= translated amino acid sequence
; NAME/KEY: misc feature
; LOCATION: (33780)...(35027)
; OTHER INFORMATION: megAI, KS5
; NAME/KEY: misc feature
; LOCATION: (35385)...(36419)
; OTHER INFORMATION: megAI, AT5
; NAME/KEY: misc feature
; LOCATION: (37068)...(37604)
; OTHER INFORMATION: megAI, KR5
; NAME/KEY: misc feature
; LOCATION: (37860)...(38120)
; OTHER INFORMATION: megAI, ACP5
; NAME/KEY: misc feature
; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAI, KS6
; NAME/KEY: misc feature
; LOCATION: (39795)...(40811)
; OTHER INFORMATION: megAI, AT6
; NAME/KEY: misc feature

; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAI, KR6
; NAME/KEY: misc feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAI, ACP6
; NAME/KEY: misc feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAI, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megAI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megAI, desosaminyl transferase, desosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megAI, TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megAI, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megAI, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence
; US-09-679-279-1

Query Match 4.1%; Score 50; DB 4; Length 47981;
Best Local Similarity 48.0%; Pred. No. 7.4e-05;
Matches 143; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 689 TGCTTTAGCGCAACGGACCGGCTGATAGACCGCTGGCACTGATGATCTACCATTTGC 748
|||||
Db 47913 TGCTGCTGGCGGATTTCGAGCGCTCGGTGAGCTCATCGGATCGGCACCTACCTGCTGC 47854
|||||
QY 749 TCAACAATCCTGAGCAGATGATGATTTTGGCTGACCGTTGCTAGTTCGAGAGCCA 808
|||||
Db 47853 TGACCCACCCGACCACTCGCGTGGTCCGGAAGACCCCGCACTGCTGCCCGGAGCGG 47794
|||||
QY 809 TTGCGGAGACATTCGTTTATAAACCGCGGTTTCAGCTGATTCGCGGAGCTGTCCCAAG 868
|||||
Db 47793 TGGAGGAGATCTGCGCTACCGAGGCCACCCGAGACACACCCGGTTCGCCACCGCGG 47734
|||||
QY 869 ATACAGTGGTGGCGGTATGGAATCAAAAGATACGATTTGTTTGTATGATCGGTG 928
|||||
Db 47733 AGGTGAGATCGCGGGGTGACGATCCCGCGGTGATGATCGACACCGCTCTGATCGCAACGGG 47674
|||||
QY 929 CGGCTAACCGGACCCCTGAAGCATTTGAACGCTGACGTGTTTAAATTCATCGGA 986
|||||
Db 47673 CGGCCAACCGGACCCCGGACAGTTTCCCGACCCCGGACCGGTTCCGACGTCACCAAGGA 47616
|||||

RESULT 6
US-09-029-603-4
; Sequence 4, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Roupel, Christine
; APPLICANT: Pospiech, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; CURRENT FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: PCT/EP96/03643
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4

```
; LENGTH: 6085
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (378)..(1665)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (1747)..(2553)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (2593)..(4011)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (4013)..(4999)
; OTHER INFORMATION: ORF
; FEATURE:
; NAME/KEY: misc RNA
; LOCATION: (5071)..(6085)
; OTHER INFORMATION: ORF
; US-09-029-603-4

Query Match      4.1%; Score 49.6; DB 3; Length 6085;
Best Local Similarity 44.5%; Pred. No. 2.9e-05;
Matches 196; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 546 CGACAGCTTCCCAATGATGTCGCGGTATTAAGACGTCGCGTCAATCCGGGATC 605
DB 989 CCAGGAGTTCACCGGTACTTCGCGGAGAGTGGACCGCGCGGCGGACGACGCGA 1048
QY 606 AGATTAAATTTGATCCTATGTAATCTCTGAATATGAAGCATGCGCGCTCGGACAAAGA 665
DB 1049 CGATCTGCTACCCCTCTCGTCCGCGCCGGGACACCGGATCACCGCTCAGGTCGACGG 1108
QY 666 TATATCGCACTGATTTAAATGTGTGTAGCCGCAACGGAACCGGCTGTAAAGACGCT 725
DB 1109 CATGTCGCGCACTGCGTCAATGCTCACCGCGGCGGACGAGACACCACTGCTCT 1168
QY 726 GCACTGATGATCACTTCTCAATCTCTGACGATGATGATGATGATGATGATGATGAT 785
DB 1169 CCGCAGGCGGCTTCTCACCTGCGCGCCACCTGACGCTCTCGAGAGCTGCGACAC 1228
QY 786 CGCTTCGTTAGTTCGAGAGCAATTCGCGAGACATTCGCGTATAAACCGCGGTTTCAGCT 845
DB 1229 ACGGAGTCGACACCGCGCGCTGGAAGAGCTGATGCGGTACGACCCCGCGCTGACGC 1288
QY 846 GATTCGCGGCGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAAGATAC 905
DB 1289 GGTGACGCGCTGGCGGTAGAGACATCCGCTCGCGGACACGACATCCCGCGGCGAG 1348
QY 906 GATTGTTTTTGTATGATCGGTGCGCTAACCGGACCCCTGAAGCATTTGAACAGCTGA 965
DB 1349 CCGGTTGTCGCGCTGCTGCGGTGCGGCAACCGGACCGCGCGCTTCCCGCTCCCGA 1408
QY 966 CGTGTTTAATATTATTCGCGG 985
DB 1409 CGTGTGAGCTCCACCGCG 1428

RESULT 7
US-09-724-797-35
; Sequence 35, Application US/09724797
; Patent No. 6733998
; GENERAL INFORMATION:
; APPLICANT: Jon S. THORSON
; TITLE OF INVENTION: MICROMONOSPORA ECHINOSPORA GENES
; TITLE OF INVENTION: ENCODING FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: CALICHEAMICIN AND SELF-RESISTANCE THERETO
; FILE REFERENCE: 2653-40
; CURRENT APPLICATION NUMBER: US/09/724,797
```

```
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 60/111,325
; PRIOR FILING DATE: 1998-12-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1209)
; US-09-724-797-35

Query Match      3.9%; Score 47.6; DB 4; Length 1209;
Best Local Similarity 53.8%; Pred. No. 5e-05;
Matches 98; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 805 GCATTGGGAGACATTGCGTTATAACCGCGGTTACGCTGATTCGGCGGAGCTGTC 864
DB 820 GCCGTGAGGAGTGCTGCGGTACGACCGCTGCGTCCAAACACCCGCGAGCTCGAC 879
QY 865 CAAGATACAGTGTGCGCGGTATGGAATCAAAAAAGATACGATTCGTTTTTGTATGATC 924
DB 880 GTCGAGCTGAGCTGCGCGGTGCGCGGTGCGCGGACGACGTCGTGCTGCTGCGGC 939
QY 925 GGTGCGGCTAACCGGACCCCTGAAGCATTTGAACAGCCTTGAGCTGTTTAAATATTCATCGG 984
DB 940 GCGCGGCGAACCGGACCCCGCGGTCGACGACCGCGCGGACGATTCGACATCGAGCGG 999
QY 985 GA 986
DB 1000 GA 1001

RESULT 8
US-08-676-818-1
; Sequence 1, Application US/08676818
; Patent No. 60571136
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Pero, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,818
; FILING DATE: 08-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/239,430
; FILING DATE: May 6, 1994
; APPLICATION NUMBER: 08/084,709
; FILING DATE: June 25, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 04599/004001
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8478
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-678-818-1

Query Match 3.7%; Score 44.6; DB 3; Length 8478;
Best Local Similarity 49.4%; Pred. No. 0.0015;
Matches 116; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 744 TTGCTCAACATCTGACGACGATGAATGATGTTTGGCTGACCGTTCGTTAGTCCGAG 803
DB 6233 TCTGCTGCAGCATCCAGAACAGCTTTTGAACCTGAGAGAAATCCAGATCTTATTGGTAC 6292

QY 804 AGCCATTGGGAGACATTCGGTTATAAACCGCGGTTCAGCTGATTCGCGGCAGCTGTC 863
DB 6293 TCTGCTGCAGCATCCAGAACAGCTTTTGAACCTGAGAGAAATCCAGATCTTATTGGTAC 6292

QY 864 CCAAGATACAGTGGTGGCGGTATGGAATCAAAAAGATACGATTTTGGTATGAT 923
DB 6353 AGAGGATATTGACATCTCGGGGTGACGATCCGTCAAGGAGAACAGTCTATCTTTTGT 6412

QY 924 CGTGGCGGTAAACCGGACCTGAGCATTTGAAACGCTGACGTTGTTTAAATTT 978
DB 6413 AGAGCGGCTAATCGAGACCTAGCATATTTCAGAACCCCGATGTTTCGATATT 6467

RESULT 9

US-09-407-549-1
Sequence 1, Application US/09407549
Patent No. 6303377
GENERAL INFORMATION:
APPLICANT: Bower, Stanley Grant
APPLICANT: Perkins, John B.
APPLICANT: Yocum, R. Rogers
APPLICANT: Pero, Janice G.
TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
TITLE OF INVENTION: SUBTILIS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/407,549
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/239,430
FILING DATE: May 6, 1994
APPLICATION NUMBER: 08/084,709
FILING DATE: June 25, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 04599/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8478
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-407-549-1

Query Match 3.7%; Score 44.6; DB 3; Length 8478;
Best Local Similarity 49.4%; Pred. No. 0.0015;
Matches 116; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 744 TTGCTCAACATCTGACGACGATGAATGATGTTTGGCTGACCGTTCGTTAGTCCGAG 803
DB 6233 TCTGCTGCAGCATCCAGAACAGCTTTTGAACCTGAGAGAAATCCAGATCTTATTGGTAC 6292

QY 804 AGCCATTGGGAGACATTCGGTTATAAACCGCGGTTCAGCTGATTCGCGGCAGCTGTC 863
DB 6293 CGCAGTCGAGGAATGTTTACGCTATGAAAGCCCCACGCAAAATGACAGCCAGAGTTGCGTC 6352

QY 864 CCAAGATACAGTGGTGGCGGTATGGAATCAAAAAGATACGATTTTGGTATGAT 923
DB 6353 AGAGGATATTGACATCTCGGGGTGACGATCCGTCAAGGAGAACAGTCTATCTTTTGT 6412

QY 924 CGTGGCGGTAAACCGGACCTGAGCATTTGAAACGCTGACGTTGTTTAAATTT 978
DB 6413 AGAGCGGCTAATCGAGACCTAGCATATTTCAGAACCCCGATGTTTCGATATT 6467

RESULT 10

US-09-105-537-21
Sequence 21, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438U1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 21
LENGTH: 1209
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-21

Query Match 3.5%; Score 43; DB 3; Length 1209;
Best Local Similarity 46.5%; Pred. No. 0.0016;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TGCTCAACATCTCTGACGACGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAG 805
DB 770 TCTTCGCGCATCCCGAGCAGTCGCGGAGCTGTGCGACCGGCCGCTCGCGGCGCG 829

QY 806 CCATTGCGGAGACATTCGGTTATAAACCGCGGTTCAGCTGATTCGCGGCAGCTGTCCC 865
DB 830 CGGTGGAGGAGACCTCTCGCTACGACCGCGCGGTGCGCTCGACGCCCGGGTGTCCGCG 889

QY 866 AGATACAGTGGTGGCGGTATGGAATCAAAAAGATACGATTTGTTTGTATGATCG 925
DB 890 GGGAGACGAGCTGGCGGCGCGGCTGCCGCGCGGCGCATGTCTGCTCTACCG 949

QY 926 GTGCGGCTAACCGGACCTCTGAAGCATTTGAACAGCTGCTGTTTAAATATTTCATCGGG 985
DB 950 CGCGACCGCGCGGACCCCGAGGCTTCACGGAACCGGAGCGCTTCGACCTCGCGCGCC 1009

QY 986 AAGATCTTGGTATCAAGAGCGCTTTTAGCGGCGCGCGCGCATCTCGCTTCGGATCC 1044
DB 1010 CCGACGCGCGCGGACCTCGCGCTGCAACCCCGCGGTCCGTACGCGCCGGTGTCC 1068

RESULT 11

US-09-105-537-3
; Sequence 3, Application US/09105537A
; Patent No. 6285202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-3

Query Match 3.5%; Score 43; DB 3; Length 13613;
Best Local Similarity 46.5%; Pred. No. 0.0068;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TGTCTCAACAATCTGAGCAGATGAATGATGTTTGGCTGACCGTTTCGTTAGTTCCGAGAG 805
Db 1575 TCCTCGCGATCCGAGCAGTGCGCGGAGCTGTGCGACCGCCCGGCTCGCGGCGCCG 1634

QY 806 CCATTGCGGAGACATTCGCTTATAAACCGCGGTTTCAGCTGATTCGCGCGCAGCTGTCCC 865
Db 1635 CGGTGAGGAGACCTCCCTCCCTACGACCCCGCGTGCAGTCGACGCGCCGGTGTCCGCG 1694

QY 866 AGATACATGTGTCGCGGTATGGAATCAAAAAAGATACGATGTTTGTATGATCG 925
Db 1695 GGGAGACGAGTGGCGGCGCGCGCTGCGCGCGCGCGCGATGTCGTCTGACCG 1754

QY 926 GTGCGCTAACCGGACCTTGAAGCATTTGAACAGCTGCTGCTGTTTATATTCATCGGG 985
Db 1755 CCGGACCGCGCGGACCGCGAGGCTTTCAGGACCGGAGCGCTTCGACCTCGCGCGCC 1814

QY 986 AAGATCTTGGTATCAAGACGCTTTTAGCGGCGCGCCCGCGCATCTCGCTTTGGATCC 1044
Db 1815 CCGACGCGCGCGACCTCGCGCTGCACCCCGCGTCCGTACGCGCGCGTGGCTCC 1873

RESULT 12

US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080

; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match 3.5%; Score 43; DB 3; Length 38506;
Best Local Similarity 46.5%; Pred. No. 0.013;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TCCTCAACAATCTGAGCAGATGAATGATGTTTGGCTGACCGTTTCGTTAGTTCCGAGAG 805
Db 35572 TCCTCGCGATCCGAGCAGTGCGCGGAGCTGTGCGACCGCGCGGCTCGCGCGCGCCG 35631

QY 806 CCATTGCGGAGACATTCGCTTATAAACCGCGGTTTCAGCTGATTCGCGCGCAGCTGTCCC 865
Db 35632 CGGTGAGGAGACCTCCCTCCCTACGACCCCGCGTGCAGCTCGACGCGCGGTGGTCCGG 35691

QY 866 AAGATACATGTGTCGCGGTATGGAATCAAAAAAGATACGATGTTTGTATGATCG 925
Db 35692 GGGAGACGAGTGGCGGCGCGCGCTGCGCGCGCGCGCATGTCGTCTCTGACCG 35751

QY 926 GTGCGCTAACCGGACCTTGAAGCATTTGAACAGCTGCTGCTGTTTATATTCATCGGG 985
Db 35752 CCGGACCGCGCGGACCGCGAGGCTTTCAGGACCGCGAGCGCTTCGACCTCGCGCGCC 35811

QY 986 AAGATCTTGGTATCAAGACGCTTTTAGCGGCGCGCCCGCGCATCTCGCTTTGGATCC 1044
Db 35812 CCGACGCGCGCGACCTCGCGCTGCACCCCGCGTCCGTACGCGCGCGTGGCGTCC 35870

RESULT 13

US-09-141-908-1
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1

Query Match 3.5%; Score 43; DB 4; Length 38506;
Best Local Similarity 46.5%; Pred. No. 0.013;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 746 TCCTCAACAATCTGAGCAGATGAATGATGTTTGGCTGACCGTTTCGTTAGTTCCGAGAG 805
Db 35572 TCCTCGCGATCCGAGCAGTGCGCGGAGCTGTGCGACCGCGCGGCTCGCGCGCGCCG 35631

QY 806 CCATTGCGGAGACATTGCGTTATAAACCGCGGTTTCCAGTGTATTCGCGGAGCTGTCTCC 865
Db 35632 CGGTGAGGAGACCTCTCGCTACGACCCGCGGTGAGTCTCGAGCGCCGCGGTGTCGGG 35691
QY 866 AAGATACAGTGTGCGGATGAGAAATCAAAAAGATACGATGTTGTTTGTATGATCG 925
Db 35692 GGGAGACGAGCTGCGGCGCGGCTGCGCGCGCGGATGTCGTCTGATGACG 35751
QY 926 GTGCGCTAACCGGACCTGAGCATTTGAACAGCTCAGCTGTTTATATTCATCTCGG 985
Db 35752 CGCGACCGCGGACCGCGGAGGTCTTACGGACCCGAGCGCTTCGACCTCGCGCGCC 35811
QY 986 AAGATCTTGGTATCAAGACGCTTTTGTAGCGGCGCGCGGCACTCTCGCTTTCGGATCC 1044
Db 35812 CCGACGCGCGGACCTCGGCTGCACCCGCGGTCGTACGCGCGGTGGCGTCC 35870

RESULT 14

US-09-657-440-19
; Sequence 19, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-657-440-19

Query Match 3.5%; Score 43; DB 4; Length 38506;
Best Local Similarity 46.5%; Pred. No. 0.013;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 746 TGCTCAACAATCTCGACGAGATGAATGATGTTTGGCTGACCGTTGTTAGTTCCGAGAG 805
Db 35572 TCCTCGGCATCCGAGCAGTGGCGGAGCTGTGCGACCGCGCGGCTCGCGGCGCG 35631
QY 806 CCAITGCGGAGACATTGCGTTATAAACCGCGGTTTCCAGTGTATTCGCGGAGCTGTCTCC 865
Db 35632 CGGTGAGGAGACCTCTCGCTACGACCCGCGGTCGAGCTCGACGCGCGGTCGCGG 35691
QY 866 AAGATACAGTGTGCGCGTATGGAATCAAAAAGATACGATGTTTGTATGATCG 925
Db 35692 GGGAGACGAGCTGCGGCGCGGCTGCGCGCGCGGCTGCGGCGCGGCGGATGTCGTCTGACCG 35751
QY 926 GTGCGGCTAACCGGACCTGAGCATTTGAACAGCTCAGCTGTTTATATTCATCTCGG 985
Db 35752 CGCGACCGCGCGGACCGCGGAGTCTTACGGACCCGAGCGCTTCGACCTCGCGCGCC 35811
QY 986 AAGATCTTGGTATCAAGACGCTTTTGTAGCGGCGCGCGGCACTCTCGCTTTCGGATCC 1044
Db 35812 CCGACGCGCGCGGACCTCGGCTGCACCCGCGGTCGTACGCGCGGTCCGTACGCGCGGTGGCGTCC 35870

RESULT 15

US-08-576-626A-1
; Sequence 1, Application US/08576626A
; Patent No. 5998194
; GENERAL INFORMATION:

; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-576-626A-1

Query Match 3.5%; Score 42.4; DB 2; Length 3756;
Best Local Similarity 47.1%; Pred. No. 0.0049;
Matches 130; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 691 CTGTTAGCCGCAACCGAACCGGCTGATAGACGCTGGCACTGATGATCTACCATTTGCTC 750
Db 648 CTGTTGCGCGGCGCGAGATGACCGCAACCGTGGTCGACGCGTCTGCGCGTCTCG 707
QY 751 AACATCTCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGCCATT 810
Db 708 GCCGAACCGGGGCTGGCCGAAACGGATCGCCGACGACCCCGCGCGCGAGCGAACCGTC 767
QY 811 GCGGAGACATTGCGGTTATAAACCGCGGTTTACGCTGATTCCCGCGCAGCTGTCCCAAGAT 870
Db 768 GCGGAGGTCGTGCGCTGCAACCGGCACTTGACCTGGAGCGCGCAGCGCCACCGCAGAG 827
QY 871 ACAGTGTGCGCGGATGGAATAATCAAAAAGATACGATGTTTGTATGATCGGTGCG 930
Db 828 GTGCGGCTCGCGAGCAGCTGATCGCGAAGCGGAGGAGTCTGTGTCGTCTGCGCGCG 887
QY 931 GCTAACCGGACCTTGAAGCATTTGAACAGCTTGAC 966
Db 888 GCCAACCGCGACCGCGAGGTCTTCCGCGAGCCCGAC 923

Search completed: December 10, 2004, 18:17:12
JOB time : 159.858 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 11:49:51 ; Search time 5403.29 Seconds
(without alignments)
10633.698 Million cell updates/sec

Title: US-10-627-124-1
Perfect score: 1215
Sequence: 1 atgagccaatgattaaatt.....tcgttgcttgtaacggggca 1215

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query		Match		Length		DB		ID		Description	
Result No.	Score	Match	Length	DB	ID	Score	Match	Length	DB	ID	Score	Match	Length	DB	ID
1	1215	100.0	47739	1	AF017113	100.0	100.0	47739	1	AF017113	100.0	100.0	47739	1	AF017113
2	1215	100.0	194692	1	BSUB0018	100.0	100.0	194692	1	BSUB0018	100.0	100.0	194692	1	BSUB0018
3	617.2	50.8	1208	6	AX433766	50.8	50.8	1208	6	AX433766	50.8	50.8	1208	6	AX433766
4	247.4	20.4	349483	1	BX571859	20.4	20.4	349483	1	BX571859	20.4	20.4	349483	1	BX571859
5	247.4	20.4	349980	6	AX770907	20.4	20.4	349980	6	AX770907	20.4	20.4	349980	6	AX770907
6	171.8	14.1	36888	1	AY260760	14.1	14.1	36888	1	AY260760	14.1	14.1	36888	1	AY260760
7	169	13.9	314100	1	SC0939106	13.9	13.9	314100	1	SC0939106	13.9	13.9	314100	1	SC0939106
8	116.2	9.6	300885	1	AE012559	9.6	9.6	300885	1	AE012559	9.6	9.6	300885	1	AE012559
9	113.6	9.3	10092	1	AE003889	9.3	9.3	10092	1	AE003889	9.3	9.3	10092	1	AE003889
10	90.8	7.5	300090	1	AE017008	7.5	7.5	300090	1	AE017008	7.5	7.5	300090	1	AE017008
11	89.4	7.4	10261	1	AE003887	7.4	7.4	10261	1	AE003887	7.4	7.4	10261	1	AE003887
12	88.8	7.3	190050	1	AL646080	7.3	7.3	190050	1	AL646080	7.3	7.3	190050	1	AL646080
13	86.8	7.1	300885	1	AE012559	7.1	7.1	300885	1	AE012559	7.1	7.1	300885	1	AE012559
14	85.2	7.0	302325	1	AE017236	7.0	7.0	302325	1	AE017236	7.0	7.0	302325	1	AE017236
15	78.8	6.5	35739	1	AF015825	6.5	6.5	35739	1	AF015825	6.5	6.5	35739	1	AF015825
16	78.8	6.5	201241	1	BSUB0007	6.5	6.5	201241	1	BSUB0007	6.5	6.5	201241	1	BSUB0007
17	75.2	6.2	110000	1	CR543861.15	6.2	6.2	110000	1	CR543861.15	6.2	6.2	110000	1	CR543861.15
18	72.4	6.0	966	6	AX433775	6.0	6.0	966	6	AX433775	6.0	6.0	966	6	AX433775
19	69.4	5.7	39228	1	MLCB1788	5.7	5.7	39228	1	MLCB1788	5.7	5.7	39228	1	MLCB1788

C	20	69.4	5.7	342300	1	MLEPRTN8	AL583924	Mycobacter
							AF274704	Pseudomon
C	21	68.2	5.6	9180	1	AF274704	Continuation (40 o	
C	22	67.8	5.6	110000	1	AE000516.39	BX842583	Mycobacter
C	23	67.8	5.6	349606	1	BX842583	AP005039	Streptomy
C	24	67.6	5.6	299925	1	AP005039	M24523 B-subtilis	
C	25	67.2	5.5	3187	1	BACRTP	AE017228	Mycobacter
C	26	67.2	5.5	300327	1	AE017228	AE017006	Bacillus
C	27	67.2	5.5	300893	1	AE017006	AX433767	Sequence
C	28	64.8	5.3	1213	6	AX433767	AE017228	Mycobacter
C	29	64	5.3	300327	1	AE017228	BX248346	Mycobacter
C	30	63.2	5.2	316050	1	BX248346	Continuation (9 of	
C	31	61.8	5.1	110000	1	AE000516.08	BX248336	Mycobacter
C	32	61.8	5.1	320050	1	BX248336	BX842574	Mycobacter
C	33	61.8	5.1	349564	1	BX842574	AF329849	Agrobacter
C	34	61.2	5.0	32495	1	AF329849	AY328003	Symbiont
C	35	61.2	5.0	75778	1	AY328003821	AE017235	Mycobacter
C	36	59.6	4.9	309267	1	AE017235	AF119621	Pseudomon
C	37	59.4	4.9	26386	1	AF119621	AB071405	Lechevali
C	38	59.2	4.9	17512	1	AB071405	AJ414559	Saccharot
C	39	59.2	4.9	25881	1	SAB414559	AB090952	Lechevali
C	40	59.2	4.9	26144	1	AB090952	AF534707	Lechevali
C	41	59.2	4.9	28654	1	AF534707	U82823	Saccharopol
C	42	58.4	4.8	8407	1	SEU82823	AE017234	Mycobacter
C	43	58.4	4.8	313846	1	AE017234	AF440524	Pseudomon
C	44	58.2	4.8	128136	1	AF440524	AE017236	Mycobacter
C	45	58.2	4.8	302325	1	AE017236		

ALIGNMENTS

RESULT 1	AF017113	Bacillus subtilis 300-304 degree genomic sequence.	47739 bp	DNA	linear	BCT 23-MAY-2000
LOCUS	AF017113	Bacillus subtilis 300-304 degree genomic sequence.				
DEFINITION	AF017113	Bacillus subtilis 300-304 degree genomic sequence.				
ACCESSION	AF017113.1	GI:2618830				
VERSION	AF017113.1	GI:2618830				
KEYWORDS						
SOURCE						
ORGANISM						
Bacillus subtilis						
Bacillus subtilis						
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.						
REFERENCE	1 (bases 1 to 47739)					
AUTHORS	Reizer, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stulke, J., Karamata, D., Sailer, M. H. Jr. and Hillen, W.					
TITLE	A novel protein kinase that controls carbon catabolite repression in bacteria					
JOURNAL	Mol. Microbiol. 27 (6), 1157-1169 (1998)					
MEDLINE	98230327					
PUBMED	9570401					
REFERENCE	2 (bases 30303 to 37035)					
AUTHORS	Robinson, C., Rivolta, C., Karamata, D. and Moir, A.					
TITLE	The product of the yvoC (garF) gene of Bacillus subtilis is required for spore germination					
JOURNAL	Microbiology 144 (Pt 11), 3105-3109 (1998)					
MEDLINE	99061203					
PUBMED	9846746					
REFERENCE	3 (bases 1 to 47739)					
AUTHORS	Lazarevic, V., Soldo, B., Rivolta, C., Reynolds, S., Maue, C. and Karamata, D.					
TITLE	Nucleotide sequence of the 300-304 chromosomal segment of Bacillus subtilis					
JOURNAL	Unpublished					
REFERENCE	4 (bases 1 to 47739)					
AUTHORS	Lazarevic, V.					
TITLE	Direct Submission					
JOURNAL	Submitted (05-AUG-1997) IGBM, Cesar-Roux 19, Lausanne 1005, Switzerland					
FEATURES						
source						
1. .47739						
Location/Qualifiers						
/organism="Bacillus subtilis"						
/mol_type="genomic DNA"						
/db_xref="taxon:1423"						
/map="300-304 degrees"						

gene	71.. .1184 /gene="prfB"	gene	ADEPTGNLDPDTSWEVMKLTLEEINNRTGTVVWATHNKEIVNTMKRKRVAIEDGIIVRD
RBS	71.. .77 /gene="prfB"	RBS	ESRGEYGSYD"
CDS	join(83..154,156..1184) /gene="prfB" /note="at low concentrations of RF-2 nucleotides 155..158 (UGAC) are translated as Asp, Phe, H.J. Nucleic Acids Res. 20, 4423-4428, 1992" /codon_start=1 /transl_table=11 /product="putative peptide chain release factor RF-2" /protein_id="AAC67303.1" /db_xref="GI:2618874" /translation="WELSIKRAELENWASRLADFRGSLDLESKEARIAELDEQMAADPE FNDQQAQTVINEANGLKDYNNYSKKLNESEHEEQMTHDLKSEPDLDQLESEKEL KSLTKEFNEFEIQLLSEPYDKNNAILLEHPGAGGTESQDWSMLLRMYTRGERRGF KYETLDYLPFGDAGIKSVTLIKGHNAVGYLKAEGVHRLVRIISPDSSGRHRTSFVS CBWMPFENDEIDIDIRTEDIKVDTYRASGAGGQHVNTDASVRIIHLPTNVVVTQOTE RSQIKNREKAMKAKLYORRIERQQAELDIRGEQKEIGWGSQIRSVVFPHPYSVK DHRVTYMGNVQAVMDGDDITFDIAYLSKLS"	CDS	3476..4385 /gene="ftsX" 3476..3484 /gene="ftsX" 3495..4385 /gene="ftsX" /codon_start=1 /transl_table=11 /product="cell division protein" /protein_id="AAC67264.1" /db_xref="GI:2618835" /translation="MIKILGRHLRESFKSLGRNTWMTFASISAVTVTLILVGVFLVIM LNLNMATNAKQVEIKVLIDLTADQKADQLONDIKELKGIQSVTFSSKEKELQQLV DSGDSGSLTKMQDENPDAFVVKVTTDPDHPDNVAKKIEKMDHVYKVYQKEEVSRLFKVGVSRNRTGIALIIGLFTAMP LI SNTIKITIFARRKEIEIMKLVGATWFWIRWPFLEGLGLGVGSVPIPIALVLSITQYVIGVWVPVKVQGSFVSLLPYNPFVQVSLVLIAGVIGWGSITSIRKPLRV"
RBS	143.. .149 /gene="prfB"	RBS	4511..5967 /gene="yvJB"
gene	1338..2195 /gene="yvJA"	RBS	4511..4516 /gene="yvJB"
RBS	1338..1342 /gene="yvJA"	CDS	4525..5967 /gene="yvJB"
CDS	1350..2195 /gene="yvJA" /note="similar to thermophilic bacterium PS-3 hypothetical protein: PIR Accession Number S43727 and to Bacillus subtilis YqfU, YpJc and YitB proteins" /codon_start=1 /transl_table=11 /product="YvJA" /protein_id="AAC67260.1" /db_xref="GI:2618831" /translation="WDVRNKTLLIIRDVYVILIGAITAVSFNVFLPNKIAAGCVSG ISTILQSGFEAAVQVMIINIPEFIAGVILLGKFGKLTAGSFLPLVFLVFTRIQPAATHLELLAIFGVGIGIVILGKSGTGTGTAQAQIIHKYSGLSLGLCIAIIDGM IYVTAMVNIHQIYAMLGVYVSKTIDVQVGFNRKSMALIITKQBOAVKEAVLQK IDRGVTKISAVGGYTDORPILMVCVVGQTEFTKLQIVKQIDESA FVI VADASEVLGE GFKA"	CDS	5967 /transl_table=11 /product="putative protease" /protein_id="AAC67263.1" /db_xref="GI:2618834" /translation="MNOKIMAVIAAGSMFGAGVYAGINLLEMDKPOTAAVPATAQA DSEEDKAMDKEKAYELISNEYVEKVDPEKLEGAIOGLMLSTLNDPYSVYMDKOTAKO PSDLSDSFEIGIABVCHMEDGKIIIVSPFKSPAKAGLKNDEILISINGSMAKD L NHAULIKRGKSSVSMKIQRPGTKKQSFRIKRAEIPLETVFASEKKVQGHSGVITIA ISTFSEHTAEDFAKALREKEIEGLVIDVRGNPGGYLQSVVEELKHFVTQDQPIYIQ IAERNGDKRYFSTLTHKKYAPVNVITDKGSASASEILGALKEAGHDVVGDTDFGK GTVQOAVPMGDGSKI LTYKWLTPNGNWIHKGIEPTIAIKQDPYFSAAGPLQKLEPL KVDNNREDVHAQVLLKGLSPDRGEDGYESDKMKKAVMAFQDQNKLNKTGVIDTRTA ETLNQOIEKKSKDENLQQLTALKSLFVN"
gene	2223..2582 /gene="cccB"	gene	6773..7977 /gene="yvJD"
RBS	2223..2229 /gene="cccB"	RBS	6773..6777 /gene="yvJD"
CDS	2244..2582 /gene="cccB" /note="putative lipoprotein" /codon_start=1 /evidence="experimental" /transl_table=11 /product="cytochrome c-551" /protein_id="AAC67261.1" /db_xref="GI:2618832" /translation="MKSLSILMIGFALS VLLAAGCSNDAAKEKTDTSKTEATASEG BELYQOSCVGCHGKDLGVSPNLQVGGKYDEHKIESIINKNGRGNPKGLVDDNEAA VIKWLSEKK"	CDS	6784..7977 /gene="yvJB" /note="transmembrane protein" /codon_start=1 /transl_table=11 /product="YvJD" /protein_id="AAC67265.1" /db_xref="GI:2618836" /translation="MSVQVQGIELLSAGLFFHLFWFFIITLAFGVVRIKREKRTF HTRIADYDDLKFTYTKGLIPEGLLSVLGIGISIPGLLAIIVATAAATLIRAN WMSAAYIVSVMLIGFGLIYQAEPLFLERFPQGFVAVMPAVAFVGLGLIITTEGAVYR SAHVRTSPALVVSRRGLPIGQOLANRWLLPLFLVPGNGLESHLSWMPVFTVPGGSF HELMIPVEFGVQVQSGLPETSIRITAKRVCIIGLAVNLGAVALSLMTPLAGAAVCT ALLGRIFLSIKQRVNDNAPFPYFKRDQGLMVLGIIPNTPAEDLELKGIEITKNGI PVKVSVDYEALQHNRAVYKLEIIGLNGEIRFDQKASYEGEHHEHGLILFVKDDRDEEA VASG"
gene	2806..3502 /gene="ftsE"	gene	complement(8025..9373) /gene="yvKA"
RBS	2806..2813 /gene="ftsE"	CDS	complement(8025..9359) /gene="yvKA" /note="belongs to the major facilitator family" /codon_start=1 /transl_table=11 /product="YvKA" /protein_id="AAC67266.1" /db_xref="GI:2618837" /translation="MSSRKWALIVSLLGAILVPIINMTIAVALSSISHTYNESIAS ITWVTVLIYMAVTPQIAGKIDMYGNKTYLWGVGLFLIASLGCALSPSLLIIVF RALQVAGALLTPNSIAIRHVSEKRLPKVFGFGLGAGLGAALGPTFGISLLIDFS WHSIFWNIPFLAIFALTALTFWTFQYKENKSDAPLDIIGSLLAGSIVSIILLTKNEA

RBS		PMGYTVSVLILFLVPLFRPRBKRTQHPILIDFALFKSFTTNANISLVLLSNLMWAVL LIMPLMTQFGLNSGMSVFSIFMSASNVWGAQJLHWRKAGKIIIFLSPAMWAG ANLLFLILSSHSVLFMLSLILGLASVGLTSMVSLATVDPGMSVAGSIFSTF RVFGSIFSSALIGLSGVHTLFMILFAVSIIGVFSIGIKSDETARIENKSA" complement (9366.. .9373)	
gene		/gene="yvkA" complement (9390.. .9972) /gene="yvkB" /note="9959" complement (9390.. .9959) /gene="yvkB" /note="belongs to the TetR/AcrR family of transcriptional regulators" /codon_start=1	
CDS		Query Match 100.0%; DB 1; Length 47739; Best Local Similarity 100.0%; Pred. No. 0; Matches 1215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGAGCCAAATCGATTAAATTTGTTAGTGTCTTTCTGTATCAATTTCAAAACAATCCATAT	60
DB	24518	ATGAGCCAAATCGATTAAATTTGTTAGTGTCTTTCTGTATCAATTTCAAAACAATCCATAT	24577
QY	61	GCTTATTTTCAACTCGCGGAGGAAGATCCGGTTTCAATTTGAAGAGTCGATAGACAGT	120
DB	24578	GCTTATTTTCAACTCGCGGAGGAAGATCCGGTTTCAATTTGAAGAGTCGATAGACAGT	24637
QY	121	TATTTTATCAGCGGCTATCATGATGTCGCTATATCTTTTCAAGATCCGATATCTTTACG	180
DB	24638	TATTTTATCAGCGGCTATCATGATGTCGCTATATCTTTTCAAGATCCGATATCTTTACG	24697
QY	181	ACGAATCACTTGTGAGCGTCCGACACGATCATCGAGGCGCTGTGTGCGCCAAATG	240
DB	24698	ACGAATCACTTGTGAGCGTCCGACACGATCATCGAGGCGCTGTGTGCGCCAAATG	24757
QY	241	CATGGAAGAAGACACTCTGCGAAGAAATTTAGTGAAGAGCTTTATCGGTGACGCA	300
DB	24758	CATGGAAGAAGACACTCTGCGAAGAAATTTAGTGAAGAGCTTTATCGGTGACGCA	24817
QY	301	CTGGATCATCTGTCTCATTTGATTAACAAATGACAGAAATCTGTGAGCGCTTATCTT	360
DB	24818	CTGGATCATCTGTCTCATTTGATTAACAAATGACAGAAATCTGTGAGCGCTTATCTT	24877
QY	361	GAAAGAGGAAAGATGATCTGTCATGATTTTGGAAAGACGTTTCCGGTGTGCGTCACG	420
DB	24878	GAAAGAGGAAAGATGATCTGTCATGATTTTGGAAAGACGTTTCCGGTGTGCGTCACG	24937
QY	421	ATGACATGCTCGGCTGGATAAAGACACATGAAATTTCTGAGTGGACACAGCGGA	480
DB	24938	ATGACATGCTCGGCTGGATAAAGACACATGAAATTTCTGAGTGGACACAGCGGA	24997
QY	481	GTTCGGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGACCATTCGTTATCG	540
DB	24998	GTTCGGATTTTATCAGAGTATCTCTCAATCTCTGAGCGCGGACCATTCGTTATCG	25057
QY	541	TGCAGCGAAGACAGCTTTCCCAATCTGATGCGCGTCAATTAAGAAAGCTCGCGCAATCCG	600
DB	25058	TGCAGCGAAGACAGCTTTCCCAATCTGATGCGCGTCAATTAAGAAAGCTCGCGCAATCCG	25117
QY	601	GGATCAGATTTAATTTTCATCTTATCTTCTGAATATGAAGGCAATCGCGTGTGGAC	660
DB	25118	GGATCAGATTTAATTTTCATCTTATCTTCTGAATATGAAGGCAATCGCGTGTGGAC	25177
QY	661	AGGATATCTCGCACTGATCTTATGCTGTGTTAGCGGCAACGAGCGGCTGATAG	720
DB	25178	AGGATATCTCGCACTGATCTTATGCTGTGTTAGCGGCAACGAGCGGCTGATAG	25237
QY	721	ACGCTGGCACTGATGATCTACCAATTTGCTCAACAACTCTGAGCAGATGAATGTTTGG	780
DB	25238	ACGCTGGCACTGATGATCTACCAATTTGCTCAACAACTCTGAGCAGATGAATGTTTGG	25297
QY	781	GCTGACCGTTCGTTAGTTCGAGAGCCATTCGCGAGACATTCGTTTATAAACCGCGGTT	840

DB	25298	GCTGACCGTTCGTTAGTTCGAGAGCCATTCGCGAGACATTCGTTATAAACCGCGGTT	25357
QY	841	CAGCTGATTCGCGGAGCTGTCCCAAGATACAGTGGTCCGCGGTATGAAATCAAAAA	900
DB	25358	CAGCTGATTCGCGGAGCTGTCCCAAGATACAGTGGTCCGCGGTATGAAATCAAAAA	25417
QY	901	GATACGATTTGTTTGTATGATTCGCGGCTAACCGGACCCCTGAAGCATTTGAACAG	960
DB	25418	GATACGATTTGTTTGTATGATTCGCGGCTAACCGGACCCCTGAAGCATTTGAACAG	25477
QY	961	CCTGACGTTGTTTAAATATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTAGCGCGC	1020
DB	25478	CCTGACGTTGTTTAAATATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTAGCGCGC	25537
QY	1021	GCCCGCATCTCGCTTTCCGATCCGCGATTCATACTGTGTAGGACGCTTTTGCAAA	1080
DB	25538	GCCCGCATCTCGCTTTCCGATCCGCGATTCATACTGTGTAGGACGCTTTTGCAAA	25597
QY	1081	AACGAAATCGAAATTTGTAGCTAATATTTGTCTGGATAGATCGGAATATCAGATTAGAG	1140
DB	25598	AACGAAATCGAAATTTGTAGCTAATATTTGTCTGGATAGATCGGAATATCAGATTAGAG	25657
QY	1141	GAAGATTTTGTATGCTGAGTCCGCTCTGTATACACGCGGACCTTTTCACTTCGTT	1200
DB	25658	GAAGATTTTGTATGCTGAGTCCGCTCTGTATACACGCGGACCTTTTCACTTCGTT	25717
QY	1201	GCCTTTGACCGGGCA 1215	
DB	25718	GCCTTTGACCGGGCA 25732	
RESULT 2			
BSUB0018/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
Bacillus subtilis subsp. subtilis str. 168			
Bacillus subtilis subsp. subtilis str. 168			
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
1 (bases 1 to 194692)			
Kunat, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,			
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,			
Borriss, R., Boursier, L., Brans, A., Braun, M., Brignall, S.C.,			
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,			
Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,			
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,			
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrarini, E.,			
Foullger, D., Fritz, C., Fujita, Y., Fuma, S., Galizzi, A.,			
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,			
Grandi, G., Guiseppe, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R.,			
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,			
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,			
Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,			
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,			
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,			
Masuda, S., Maue, C., Medigue, C., Medina, N., Wellado, R.P.,			
Mizuno, M., Moest, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,			
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,			
Portetle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,			
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,			
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,			
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,			
Sekowska, A., Seror, S.J., Serrero, P., Shin, B.S., Soldo, B.,			
Sorokin, A., Taccioni, E., Takagi, T., Takahashi, H., Takemaru, K.,			
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognoni, A.,			
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,			
Viari, A., Wambutt, R., Wedler, H., Wedler, H., Weitzenegger, T.,			
Winckers, P., Wipac, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,			
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and			
Danchin, A.			

TITLE	The complete genome sequence of the gram-positive bacterium
JOURNAL	Bacillus subtilis
MEDLINE	Nature 390 (6657), 249-256 (1997)
PUBMED	98044033
REFERENCE	9384377
AUTHORS	2 (bases 1 to 194692)
TITLE	Kunat, F., Ogaawara, N., Yoshikawa, H. and Danchin, A.
JOURNAL	Direct Submission
	Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur,
	Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724
	Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
	adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
COMMENT	68 89 48
	On Jul 7, 2003 this sequence version replaced gi:2635827.
	This entry contains data from release R16.1 of the Subtilist
	database. Further data on gene annotation and detailed information
	about changes from previous releases can be found at
	http://genolist.pasteur.fr/Subtilist/.
FEATURES	Location/Qualifiers
source	1..194692
	/organism="Bacillus subtilis subsp. subtilis str. 168"
	/mol_type="genomic DNA"
	/strain="168"
	/db_xref="taxon:224308"
gene	complement (79)..888
	/gene="fhuC"
CDS	/locus_tag="BSU33290"
	complement (79)..888
	/gene="fhuC"
	/locus_tag="BSU33290"
	/codon_start=1
	/transl_table=11
	/product="ferrichrome ABC transporter (ATP-binding
	protein)"
	/protein_id="CAB15335.1"
	/db_xref="GI:2635843"
	/db_xref="GOA:P49938"
	/db_xref="Swiss-Prot:P49938"
	/translation="WNSLSTELGIGYGDRTIVEDLNI SIPKITTLLIGNGCGKST
	ILKTSRIMRSAGAVYNGKAIHKMSKTDIAKDWAILPQPEAPSGLTVHELVSYGR
	FPHQSGFRDNDEDRRIIKWALEETGMAEYARPEALSGGQRQORVWIALAQTGL
	LLDEPTTDLAHOLEILQLDRLNKEQGRITLMIHDLNHAARFSHYMLAKKGTV
	IKEGTALEWTPDILKQVFDIAEIVTDPRTNKVCLATYDLIKNEKELTV"
gene	complement (904)..1914
	/gene="fhuG"
CDS	/locus_tag="BSU33300"
	complement (904)..1914
	/gene="fhuG"
	/locus_tag="BSU33300"
	/codon_start=1
	/transl_table=11
	/product="ferrichrome ABC transporter (permease)"
	/protein_id="CAB15336.1"
	/db_xref="GI:2635844"
	/db_xref="GOA:P49937"
	/db_xref="Swiss-Prot:P49937"
	/translation="MKTINKPLIYVAVTELLIVVFISLNLGVIKIAPLKLQVFP
	GGTARDELVFEFLPRILLISLLVAGISVAGAILQSVQNELAEPLGILNAGGSL
	AVLFIYFFQSADLSFFGTGFMPLFSALAGAILAFLIYLAWKKGVTPIRLIDGI
	YAGNFANLLILFQKMDPHDFMQAAVWISGWNANWMIWAILPWIILLFLTYK
	ARYLNMQLDGLATGLGTAVKERRILLAAVTLAASCAVAAAGGIAFLGIAPHVAR
	RLTGPRHQTLIPVSFISFLPLADTLARNVLATSEIPVGVISVLGAPFYIYLLMK
	AN"
gene	complement (1914)..3068
	/gene="fhuB"
CDS	/locus_tag="BSU33310"
	complement (1914)..3068
	/gene="fhuB"
	/locus_tag="BSU33310"
	/codon_start=1
	/transl_table=11
	/product="ferrichrome ABC transporter (permease)"
	/protein_id="CAB15337.1"
	/db_xref="GI:2635845"
	/db_xref="GOA:P49936"
	/translation="MHPFCSKHSIKSAEKSDILKQOLIIILSNRKEVQLSOHKNIR
	TASBEIQWTSRTYGAVIYIAGLCLGLGAFSLISLGAADHLRTVWEAITHYQPTKT
	SHQIHDLRLPRTAAALVAGLAVSGAIMQGMTRNPLAEPISIMGVTSGSAFIAF
	APFFCLAGMLVMSFAGAGLGAFTVMGIGMFSRGGTLPVKLALAGTAVTFFTGIST
	AIARLFDVAQDISFWAGVAGVKGWGLLLIAGAVGLTAFFARVTVLSLGGDL
	AKGLGOYTSVAKLVGMLIVLITGAAVSIAGTIAFIGLII PHITFLVGVDRWIIIC
	SAVLGAVLLVPADIAARLVNAPFTPTPGALISLIGVPPFFYLARRERGL"
gene	3166..4113
	/gene="fhuD"
CDS	/locus_tag="BSU33320"
	3166..4113
	/gene="fhuD"
	/locus_tag="BSU33320"
	/codon_start=1
	/transl_table=11
	/product="ferrichrome ABC transporter (ferrichrome-binding
	protein)"
	/protein_id="CAB15338.1"
	/db_xref="GI:2635846"
	/db_xref="GOA:P37580"
	/db_xref="Swiss-Prot:P37580"
	/translation="WTHIYKLGAAFFALLIIAALACGNNSSEKSGSADSKGAEFTT
	YKAENGIVKI PKHPRVVMADGYGVFKTLGINVVGAPENVKPNPYKGTNGVNI
	GDGTSVKVIDLNDPLIIVMTTQAGDIKKLEKIAPTVAVKYDKLDNIQLKEPAKWTG
	TEDXAERKLAKWKKVAAAKYIKKAVGDKTISIMQNGKDIYVFGKDFGRGGSIIYK
	DLGLOATKLTKKKAIDPGYTSISLEKLDPFAGDIYPAGPWQSGDDGGVFPESIWK
	NLNAVXNGHVYKMDPIGIFTDPIISLEGQLEFITESLTK"
gene	complement (4348)..5757
	/gene="yvsh"
CDS	/locus_tag="BSU33330"
	complement (4348)..5757
	/gene="yvsh"
	/locus_tag="BSU33330"
	/function="unknown"
	/note="similar to ABC transporter (amino acid permease)"
	/codon_start=1
	/transl_table=11
	/protein_id="CAB15339.1"
	/db_xref="GI:2635847"
	/db_xref="GOA:O32204"
	/db_xref="TrEMBL:O32204"
	/translation="MEQTKWGFMLLTAFVGNVVGSGIFSLPSSLASIASPFGATSA
	WLLTGAVLMIALVFGHLSIRKPELTAGPQSYARALFSDPKKGNRAGTWMGYWAS
	WISNVAITSLAGYLTSPFPLVDKREMFISGGQEVTLGQLITFAVCTILLWGHAIL
	VASINGASKLNFVTLTKVLGFFVFI VAGLFVFTSLFGHPFYFQSGNGTSIGGGQ
	VHNAISTLWAFVGSIAVILSGRARSQDVKRATITGLLIALSIYIIVTITMGVLP
	HDKLVGSEKPPVDVLYAIVGNAGSVIALLAILCLFGTLMGLWSVVPYQAAKAGD
	PPAFFAKTKKGSPIVAILITNVMSQVPIFSVISRTISDAFTFLTATLAVLIPLV
	SAIYSLKVIKGETYDQKGSVRDGLIALLACAYSVPVITGTADLTFTFILGLFF
	VGLIYPPVSNKFKQEKQA"
	complement (6157)..6297
	/gene="aspJ"
	/locus_tag="BSU33340"
	complement (6157)..6297
	/gene="aspJ"
	/locus_tag="BSU33340"
	/codon_start=1
	/transl_table=11
	/product="small acid-soluble spore protein"
	/protein_id="CAE01467.1"
	/db_xref="GI:32468826"
	/translation="MGPFNDKKGKRSEKKNVIOGALEBDAGSALKDDPLQBAVQKKKN
	NR"
gene	6464..6946
	/gene="yvsG"
CDS	/locus_tag="BSU33350"
	6464..6946
	/gene="yvsG"
	/locus_tag="BSU33350"
	/function="unknown"

QY	239	TCATCGGAAAGAACACCTCTGCCAAAGAGAAATTTAGTGTGAGAGCTTTATCGGTGACG	298	ACCESSION	BX571859	1	GI:36783455
Db	241	TGAAAGGCAAGAGACACGCGCAAAAGAGAGATCGTTCTGCGCGCTTTATCGGGGAAT	300	VERSION	BX571859	1	GI:36783455
QY	299	CACTGGATCATCTGCTCAATTGATTAACAAAATGCGAGAAACTTGTAGCGCTTATC	358	KEYWORDS	complete genome.		
Db	301	CCCTTGATCACCTCACCGCTCATCAAGAAATGCGCAAGGCT--TTTGCCCGCGCAC	358	SOURCE	Photorhabdus luminescens subsp. laumondii T101		
QY	359	TTGAAAGAGGAAAGTAGTCTCGTCAATGATTTTGGAAAGAGCTTTGCGGTGTCGTC	418	ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.		
Db	359	GTGGAGAGGGCGGATCGACCTTGTCATGATTTTCGGCAAAATTCGCGGTTTCGGTGA	418	REFERENCE	Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Taourit, S., Bocs, S., Boursaux-Bude, C., Chandelier, M., Dassa, E., Derose, R., Derzelle, S., Freyssinet, G., Gaudriault, S., Givaudan, A., Glaser, P., Medigue, C., Lanols, A., Powell, K., Siguier, P., Wingate, V., Zouine, M., Boemare, N., Danchin, A. and Kunst, F.		
QY	419	CGATGACATGCTCGCGCTGGATAAAGAGACCATCAAAAATCTCTAGTGGCACAGCG	478	AUTHORS	Complete genome sequence of the entomopathogenic bacterium Photorhabdus luminescens		
Db	419	CGATGGACATTTAGTTTGGACAGACGACCAACACGCTGAGAACTGGCACAGCG	478	TITLE	Nat. Biotechnol. 11 (1) (2003) In press		
QY	479	GAGTTGCGGATTTATCAGAGATCTCTCAATCTCTGAAGCGGGCACATTCGTTAT	538	JOURNAL	Duchaud, E., Frangeul, L., Rusniok, C. and Kunst, F.		
Db	479	CGTCGCCGATTTCAATTACAGTTTGAATCAGGCGCTGAGGACCGGAGCATTCCTCTCA	538	REFERENCE	Direct Submission		
QY	539	GGTGACGGAACAGCTTTCCCAATCTTGATCGCGTCAATTAAGAACTGCGCTCAATC	598	AUTHORS	Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: lfrangeul@pasteur.fr, fkunst@pasteur.fr		
Db	539	AATGAGTGAACAGCTCGCTGAGTATTTGAATCCGATATCGAGGAAAGCGCAAAATC	598	JOURNAL	Location/Qualifiers		
QY	599	CGGGATCAGATTTAATTTTCGATCTCTATGATCTTCTGAATATGAAGCGCATGGCGTGTGCG	658	FEATURES	1. 349483		
Db	599	COGGACATGATTAATCATCTCTTGGACATTCGACATACGAGGAGTGGCGATGTCTG	658	source	/organism="Photorhabdus luminescens subsp. laumondii T101"		
QY	659	ACAAGATATATCGGACATCTTAATGCTGTAGCCGCAACGGAACCGGCTGATA	718	gene	/mol_type="genomic DNA"		
Db	659	ACCGGACATACGCGCGCTGATCTCAATATATTTGCTTCCGCCACAGAACCGCGGACA	718	RBS	/strain="T101"		
QY	719	AGACGCTGCATGATGATCTACCATTTGCTCAACATCTCAGCAGATGATGATGTT	778	CDS	/db_xref="taxon:243265"		
Db	719	AGACGCTGCATTAATGATTTATCATTTGCTTCATCATCCCGATCAGATGAATGATGTC	778		/gene="dnaA"		
QY	779	TGCGTGACCGTTCGTTAGTTCGAGAGCCATTCGCGAGACATTTGGTTATTAACCCGCGG	838		/locus_tag="plu0001"		
Db	779	TGGAAGCCGACGCTTCTCCCTCAGGCAATCGCAGAGCGCTGCGCTACAGCCGCGG	838		/gene="dnaA"		
QY	839	TTGAGTGAATCCGCGGACGCTGTCCCAAGATACAGTGTGTCGGGTATGGAATCAAAA	898		/locus_tag="plu0001"		
Db	839	TGCAGCTCATCCGCGCCAGCTTTACAAAGACGCGGAGATTGCGGAGTTCGAGCTTAAAG	898		/codon_start=1		
QY	899	AGATACGATTTGTTTGTATGATCGGTGCGGTAAACCGGACCTGAGCATTTGAAC	958		/transl_table=11		
Db	899	AAGGACGAGCTGATTTTGCATGATAGGCGCGCAATCGCATCTGAAGCGTTCGAGG	958		/product="Chromosomal replication initiator protein DnaA"		
QY	959	AGCCTGACGTGTTTAAATATTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTTAGCGGG	1018		/protein_id="CAE12296.1"		
Db	959	ATCCCGACAAGTTCAACATTCACCGCAGCGACTTGAAGTCAAGGCGATTCAGCGGG	1018		/db_xref="GI:36783456"		
QY	1019	CGCCCGGCAATCGCTTTTCGATCGGCAATTCATACTGTGTAGAGCAGCTTTTGCA	1078		/translation="MSLSWQCLARLQDELPAEFMSWIRPQAEISGNTLALVAPN RFVLDWPREKYINNINGLNDPCGAEVPLLRFEVGNKPVSONDSPQRVVTHTPVAPA PONTVRSWNTAVQPELSYRSVNPVNPKHTFDNFVEGKSNQLARAAQVADNPGAY NPLYRGSTGTGKTHLHAGVNSIMARKANAKVVMHSERFVQDMVKALPQNEIIEFK RYYSVDALLDDIIOFFANKERSQEEFFHTFNALLEGQQIILTSDRYPKNEINBEFR LKSRFGWLTVAIEPPELSETRVAIIMKKADENEIQLGEVAFIAKRLSRNVRELEGA LNRVTANANFTGRTITIDVREALDLALQELVTIDNIOKTVAEYKIKVADLLSK RRSRVAPPRPMAMALAKELTNHSLPEIGDAFGGRDHTTVLHACKRIEQLRESHDIK EDPENLRITLSS"		
Db	1019	CAGCCAGGCAATCTCGATTCGGTTCAGGCGTCCACAACTGTGTAGAGCGGGTTTGCA	1078		/gene="dnaN"		
QY	1079	AAAGCAATCGAATTTAGCTAATATTTGCTGATAGATGCGGATATCAGATTAG	1138		/locus_tag="plu0002"		
Db	1079	AGACCGAAATGAGCTTGTGCAAAATATCTCTTATCAGCTGAAATATTCGGCTGG	1138		/codon_start=1		
QY	1139	AGGAAGATTTTGTATGCTGAGTCGGGTCTGTATACGCGGACCTGTTTCACTTCTCG	1198		/product="DNA polymerase III, beta chain"		
Db	1139	AGGAGACTTCATTATCCCGGAGACAGGGCTTTACACAGCGGACCCGCTCTCGCTCAAC	1198		/protein_id="CAE12297.1"		
QY	1199	TTGCGTTTGA	1208		/db_xref="GI:36783457"		
Db	1199	TCCGTTTGA	1208		/translation="MKFIIEEQLLKPLQOVSSPLSGRPTLPILGNLLLOVTEGSIILL TGTDLMEMMARVTLSPHEGATVPARKFDFIWRGLPDGAEISVELDGRLLYRSPG RSRFSLTPASDFNLDQSEVFSPLQATLKRLLIESTQFMAHQDVRYYLNGWLF ETEGELRTVATDGHRLAVCSGMIGQKLPKSHSVIPRKGVIELMRLLDGGDTPLQIGI GSNIRAHVGDPIFTSKLVDRGFPDYPYRVLKPNPKDLEANCMDLQAFSPRAAILNE KFGVRLYFSENQLRITANNPEQEAEEIVDVSYQGAEMEIGFNVSYVLDVLTNLRCE EYRLLLTDAISSVQIEDCANNAAYVWPMRL"		

RESULT 4
BX571859/c
LOCUS
DEFINITION Photorhabdus luminescens subsp. laumondii T101 complete genome;
BX571859 349483 bp DNA linear BCT 26-SEP-2003

QY 281 GAAGCTTTATCGGTGACGCACTGGATC---ATCTGTCTCATTGATTAACAAAATGCAG 337
Db 314612 AAGGACTTTCAGAGAGATATTTTATTCGTACTACGACCCATGATTCGAAGATTACTG 314553
QY 338 AAACTTGTTAGCGCTTATCTTGAAGAGGAAAGTGATCTCGTCAATGATTTTGGAA 397
Db 314552 AAGATCTCTTCAACACCTTATATGGAAGAAAGGAATATAGATATGTTAATGACTTTGGCC 314493
QY 398 AGACGTTTGGGTGTCGCTCAGATGGACATCTCGGGTGGATAAAGAGACCAATGAA 457
Db 314492 GCGATTATGCCGTATAGTGACATTAAGATTTCTCGGATGCCAAGTGATACTACTCGT 314433
QY 458 AAATCTCTGAGTGGCACACGCGAGTTCGCGATTTTATCACAGATATCTCTCAATCTCCTG 517
Db 314432 ATATAGCTGAATGGCACAGGGTATCGCCAGTTTATTAACCCAGTTTGACCAACAGAAC 314373
QY 518 AAGCGGGGCAATTCGTTATGGTGAGGAGCAAGCTTTCCCAATCTATGATGCGGTCA 577
Db 314372 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAAAATTAATTCGCTTACTTAAAGCCATAA 314313
QY 578 TTAAAGAGCGTCGCTCAATCCGGGATCAGATTTAATTTTCGATCCTATGACTTCTGAAT 637
Db 314312 TAGATCAGAGACGGGTATATCCAGTAAGGACATTAATCTATAATCTG----- 314264
QY 638 ATGAAGGATGCGCTGTGCGCAAGGATATATCTCGCACTGATTTCTTAATGTGCTGTAG 697
Db 314263 -TCAAGATACAGGATGTCATGAGTGAATTAACAGCACTGTGTTAAACATCTTATTAG 314205
QY 698 CGCAACGGAAACCGGTGATGAAGCGCTGGCACTGATGATCTACCAATTTGCTCAACAATC 757
Db 314204 CGGCTACAGAACTGTGCAAGAGATATTAGCAATGATGCTTAATCACTTAATCTAATC 314145
QY 758 CTGACGAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGCCATTCGGAGA 817
Db 314144 CGAGTATGCTGATGTAGTCTCAAGATTCGAGCTTAGTTCGAGATGCAATTTGAAGAA 314085
QY 818 CATTCGGTTATAAACCGCGGTTCAGCTGATTTCCGCGCAGCTGTCCCAAGATACAGTGG 877
Db 314084 CTTTAGCTTGACATCCCGCTACAACTCATTTCCAGAGAGCGGAGTGAGACGCTCACTA 314025
QY 878 TCGCGGATGAAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGGCTAACCC 937
Db 314024 TTCTGGTATCGATATACCTAAAGGCGGTAGTATTTGTATGATTTGGCGCAGCTAATC 313965
QY 938 GGGACCTGAACATTTGACAGCCTGACGTGTTTAATTTTCAATCATCGGGAAGATCTGGTA 997
Db 313964 GTGATCCATCGGTTTTTCAACAAACCAATGAATTTGATCTATATCGAAGAAAAATACCA 313905
QY 998 TCAAGAGCGCTTTTAGCGCGCGCGCGCATCTCGCTTTCCGATCCGGCATTCATAACT 1057
Db 313904 CTTCTCCAGAAAGCAATCGNAAAGACATTTAGCTTTTGGTGCAGGTACTCATGCT 313845
QY 1058 GTGTAGGACAGCTTTTGCNAAAGCAAAATCGAAATTTAGCTAATATGTCGTGGATA 1117
Db 313844 GTGCTCGGCTGCAATTTTCACTCAGTCACTTGAAGTTTCACTCCAATATTTATTTGGATC 313785
QY 1118 AGATCGGAATACGATTAGAGGAGATTTTGTATGCTGAGTCCGCTCTGTATACAC 1177
Db 313784 TCGTGATAATTTACGTTTTTGTGTCATCACTATATCAAGAAACAGCGGTGTATACAC 313725
QY 1178 GCGGACCTGTTTCACTTCTCGTTGCGTTGA 1208
Db 313724 GAGGACCTTCTAAACTCCTTTGAGTTTGA 313694

RESULT 5
AX770907
LOCUS AX770907 349980 bp DNA linear PAT 02-JUL-2003
DEFINITION Sequence 38 from Patent WO02094867.
ACCESSION AX770907
VERSION AX770907.1 GI:32438071
KEYWORDS Photorhabdus luminescens

ORGANISM Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
1
REFERENCE
AUTHORS Duchaud,E., Taurit,S., Glaser,P., Frangeul,L., Kunst,F.,
Danchin,A. and Buchrieser,C.
TITLE Sequence of the Photorhabdus luminescens strain T701 genome and
uses
JOURNAL Patent: WO 02094867-A 38 28-NOV-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1. 349980
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/db_xref="taxon:29488"
/note="sequence length to big . Split into 2
sequences.--seq 00038: 1 too 349980--seq 10787: 300001 to
618776"

ORIGIN

Query Match 20.4%; Score 247.4; DB 6; Length 349980;
Best Local Similarity 52.7%; Pred. No. 4.3e-65;
Matches 617; Conservative 0; Mismatches 536; Indels 18; Gaps 3;
QY 41 AATTTCAAAACAATCCATATGCTTATTTTTCACAACCTGCGGAGGAAGATCCGGTTCAAT 100
Db 283246 AATTCAATAAATCTTACGATTTTACGACATTTTGCATTAACAAGATTTAGTTTAT 283305
QY 101 ATGAAGATCGATAGACAGTTATTTTATCAGCGCTATCATGATCGCGTATATCCTTC 160
Db 283306 TTGAACAATCCAAATAGCTATTTTATTTGCAATATAGAGATGTGACGCTATTTAA 283365
QY 161 AGCATCCGATATCTTACGAGCAATCACTGTTGAGGTGCGCAACAGATCATCGGAG 220
Db 283366 AATCCTCA--ATTTCATATACTAAGCCATTAACATGCACTTCCGAAACGGTTATGGGG 283422
QY 221 GCCCTGCTCGCCCAATGATGGAAGAACAACCTCTGCCAAAAGAGAAATTTAGTGA 280
Db 283423 ATCGGTTCTTGCTCAAAATGAAGTGAAGCATGCGTGTAAACGAAATTCATAATGC 283482
QY 281 GAAGCTTTATCGGTGACGCATCTGGATC---ATCTGCTCCATTTGATTAACAAAATGCAG 337
Db 283483 AAGGACTTTCAGAGATATTTTAAATCGCTACTACGAACCCATGATTCGAAAGATTACTG 283542
QY 338 AAACTGTTAGCGCTTATCTTGAAGAGGAAAGTATCTCGTCAATGATTTTGGAA 397
Db 283543 AAGATCTCTTCAACCTTATATGGAAGAAAGGAATATAGATATTTGTTATGACTTTGGCC 283602
QY 398 AGACGTTTGGGTGCGTCACGATGGACATCTCGGGCTGGATAAAGAGACCAATGAA 457
Db 283603 GCGATTATGCGTATTAGTGACATTAAGTATTTCTCGGACTGCCAAGTATATACTATCGTG 283662
QY 458 AAATCTCTGAGTGGCACAGCGAGTTGCCGATTTTATCACAGATATCTCTCAATCTCCTG 517
Db 283663 ATATAGCTGAATGGCACAAAGGTATCGCAGTTTATTTACCCAGTTTGACCAACAGAAC 283722
QY 518 AAGCGGGGCAATTCGTTATGGTGACGGAACACCTTTCCCAATACTTGTATGCCGCTCA 577
Db 283723 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAATTAATTCGCTTACTTAAAGCCATAA 283782
QY 578 TTAAAGAGCGTCGCTCAATCCGGATCAGATTTTAAATTTTCGATCCTATCTACTTCTGAAT 637
Db 283783 TAGATCAGAGACGGGTAAATCCAGTAAGGACATTAATCTATAATCTG----- 283831
QY 638 ATGAAGGATGCGCTGTGCGCAAGGATATATCTCGCACTGATTTCTTAATGCTGTAG 697
Db 283832 -TCAAGATACAGCGATGTCATGAGTGAATTTACAGCACTGTGTTTAAACATCTTATTAG 283890
QY 698 CGCAACCGAACCGGCTGATGAAGCGCTGGCACTGATGATCTACCATTTTGTCTCAACAATC 757
Db 283891 CGGCTACAGAACTGCTGCAAGATATTAGCAATGATGCTTAATCACTTAATATCTAATC 283950


```
QY 758 CTGAGCAGATGAATGATGTTTGGCTGACGCTTCTAGTTCGGAGAGCCATTGGGAGA 817
Db 283951 CGAGTATGCTGATGATGTTCTCAAGATCGCAGCTTAGTTCGAGATGCATTGAGAAA 284010
QY 818 CATTGGGTTTAAACCGCGGTTTCACTGATCCGGCGGAGCTGTCCCAAGATACAGTGG 877
Db 284011 CTTTAGCTGACATCCCGTACACTCATTTCCAGAGAGCGGAGTGAAGAGCTCACTA 284070
QY 878 TCGGCGGTATGGAATCAAAAAGATACAGTATGTTTTTTTGTATGATCGGTGGCGCTAAC 937
Db 284071 TTTCTGGTATCGATATACCTAAAGCGCTGTAGTATTTTGTATGATGTTGGCGAGCTAATC 284130
QY 938 GGGACCTCAAGCATTTGAACGCTTGAAGTGTGTTTAAATATTCATCGGAGAGATCTTGTA 997
Db 284131 GTGATCCATCGGTTTTTCAAAACCAATGAATTTGATCTATCATCAAGAAAAATACCA 284190
QY 998 TCAAGAGCGCTTTTAGCGCGCGCGCGCATCTCGCTTTTCGGATCCGCGCATTCATAACT 1057
Db 284191 CTTCTCCAGAAAGCAATCGAAAGACATTTAGCTTTTGGTCAGGTACTCATGCT 284250
QY 1058 GTGTAGGACGCTTTTGGCAGAAACGAAATCGAAATGAGTAAATATGTTGCTGGATA 1117
Db 284251 GTGTCGGCTGCTATTTTCACTCAGTCAGTGAAGTTTCATCCATATATTTTGGATC 284310
QY 1118 AGATCGGATATCAGATTAGAGAGATTTTGTATGCTGATCGCTGCTGTATACAC 1177
Db 284311 TCTCGTAAATTTACGTTTTTGTGATCAGTATCATCAAGAAACAGCGGTGATACAC 284370
QY 1178 GCGGACCTGTTTCACTTCTCGTTCGCTTGA 1208
Db 284371 GAGGACCTTCAACTCCCTTTGAGTTTGA 284401
```

RESULT 6

LOCUS

DEFINITION

AY260760 16688 bp DNA linear BCT 30-APR-2003
Streptomyces hygroscopicus subsp. yingchengensis Shy1 (shy1), Shy2 (shy2), Shy3 (shy3), Shy4 (shy4), Shy5 (shy5), Shy6 (shy6), Shy7 (shy7), Shy8 (shy8), Shy9 (shy9), Shy10 (shy10), Shy11 (shy11), Shy12 (shy12), Shy13 (shy13), Shy14 (shy14), Shy15 (shy15), Shy16 (shy16), Shy17 (shy17), Shy18 (shy18), Hmr19 (hmr19), Shy20 (shy20), Shy21 (shy21), Shy22 (shy22), Shy23 (shy23), Shy24 (shy24), and Shy25 (shy25) genes, complete cds.

ACCESSION

AY260760

AY260760.1 GI:30267715

VERSION

KEYWORDS

Streptomyces hygroscopicus subsp. yingchengensis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 16688).

Qin, L., Wang, H., Wu, Z., Lu, Y. T., Deng, Z., and Zhao, G.

Identification and function analysis of a MDR efflux protein-Hmr19,

in a sequenced genomic DNA fragment from Streptomyces hygroscopicus

yingchengensis 10-22

Unpublished

2 (bases 1 to 16688).

Qin, L., Lu, Y. T., Zhao, G., Bao, K., Deng, Z., Ren, S. and Jin, W.

Direct Submission

Submitted (23-MAR-2003) Genome Laboratory, Shanghai Research Center

of Biotechnology, CAS, 500 Cao Bao Road, Shanghai 200233, China

Location/Qualifiers

1. 36688

/organism="Streptomyces hygroscopicus subsp.

yingchengensis"

/mol_type="genomic DNA"

/isolate="10-22"

/sub_species="yingchengensis"

/db_xref="taxon:228732"

/clone="pHZ1392"

983. 2209

/genes="shy1"

983. 2209

gene

CDS

```
/gene="shy1"
/feature="putative P450 protein"
/codon_start=1
/transl_table=11
/product="Shy1"
/protein_id="AAP21648.1"
/db_xref="GI:30267716"
/translation="MSRAHLLPDILSPFAANPYAYVMREKEBLIWHTEATQSYIIS
RYEDVERFVKDKABFTDNNVWQLEPVHGKTIQLQSGREHVAVRALVAPAFRGSDLE
KQFLPVIERNREKELIDAFRTGSDI VNDYATFPVNTIADMLGDKADHAFPHGMYT
AVIAPLNLGSDPEVAAGERTRVFAFVLMFVIRERRANPGDDLLSLTCAAEVDGVR
MSDEDI KAFCSLLLAAGGTTDKAIGILANLLSHPDOLAAVRADSLI PAAPAEFLR
YTPPQVIMROSATDEVTGTTI PAGATVVTCLIGANRDERRYRDPDRDI ERDDLAT
TSAPRAAGHLA PALGRHFCVGLLAKAEVEVGLNQLLDAMPDLRLADGHDILVEQGVF
TRGPXILPVRFTPTTA"
complement(2270..4150)
/gene="shy2"
/complement(2270..4150)
/gene="shy2"
/feature="putative P450 like ATP/GTP binding protein"
/codon_start=1
/transl_table=11
/product="Shy2"
/protein_id="AAP21649.1"
/db_xref="GI:30267717"
/translation="NGSATSELSORTPLTAAETCLKIVVGGVGGKTLTVRSVSE
IRPLTEEMTQAGQIDETAGVERKTTTVAFDGRISLNDMLVILFGAQGRFV
FLWDLRFAGTGAIVLVTRMEDCWAIDRLHLEHGTFFVAVNPFSDGDEKRFSDLE
ROALALGHPVMEICDVRASKEVLI ALVDHLYTRALAKESTACSDTTFGPFSDAP
PPGPAHGSAPVLAQYQOTPSQLYRLRREHGAVALVLDGGI PAWLVGPEVCY
VTADELPAHDSRRWNEHIPPDMPLPYCYQPSVLTGAEHQRRAVITQALEG
VDQELARECLIAARLISFSGSRAELMSYAHALPARGVLMCGMPAEDADTERL
VDDLRLSIDAGGDDPVAAYTRVGERIMRLVKEKRRPGPDVTSMLHPAGLGDDEI
VQDLISVIAAAQPTANMTCNTLRLLTDERFVAVNAGRVSGSALNEVLWLDPTQ
NFIGWAVRDTQLGGRHREGCLVLGAANTDQIWPHPAGSNAHLDFSNGEH
RCPYAPLLADVMARTAVETLLEHLPLDLVLAPEELTWPSIWMRGLTSLPVEPTPA
NN"
complement(5031..6773)
/gene="shy3"
/complement(5031..6773)
/gene="shy3"
/feature="putative sensor kinase"
/codon_start=1
/transl_table=11
/product="Shy3"
/protein_id="AAP21650.1"
/db_xref="GI:30267718"
/translation="MPIPARSQRRGAGGAAPASPVALPVLAVTAAGAATAVLAP
DGRVWGTGVVWNTICIAVSVRASVIVRRRAHRAEQNELELARNWRHQAETSR
LVDVILPDVQVRNGTGVVEALSTVLPSPDHLQRLKRVATELATS VGRALAESR
LEDVRAQLERTAEILRRVABETLPAIALLREGSADTVLAKFDHPHNVLLRSSAERF
VRELAYSRRSAAQAASAKALSRVQAKTSMADLREMQERHGEVFGDLRLDHSF
SOLGMDRLALLMGSRSVNMKPIVMESILRGAVRIAAYRRVRLHSSRAAISAF
RAEGVMHLLAELMDNAANFSPIDVHVYVEERSAGLVITIEDSGLKMSDAAMRAVE
SVSGRTDLASLQGTSLGLGVGVLA VKYISVSRPSRGGTGVVILPPLVAQRH
EPLTPRQAPAPHSPARPADEPLTILPEAPARTAPPAPAPAPATSDVTAGGG
ATANGLPVAPAGRAAEATRRKPAQTADPVRVFGPALDAGSRFGPHRRRAEG
ASADGDTHTDTPDA"
complement(9017..9682)
/gene="shy4"
/complement(9017..9682)
/feature="putative transcriptional regulatory protein"
/codon_start=1
/transl_table=11
/product="Shy4"
/protein_id="AAP21651.1"
/db_xref="GI:30267719"
/translation="MTVAGKETPRERYRAQLKAEIKERAWEQIAAAGATGLSLNIAIK
QMMSGPALRYFASRDELITELIRDAYSSLDITRTAAAGADLAVLGRALRAWALD
DPQRYELIYGPVGVYHAPDDITATASEIMOALLDAAOPTKTVTVTDGHPDTRLEA
HLAQHWAAGHPAPPAALRRALLFWTRLHGILSLELAGHFTGMGLDPAELVDELRLH
LSQ"
```



```
QY 829 AAACGCGCGTTCAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGTGTCGGCGGTATG 888
Db 1814 ACACGCGCGTCCAGATGATCATGCGGCGAGTCCGGCCACCGAGGTCAACCGTGGC 1873
QY 889 GAAATCAAAAAGATACGATCTTTTTCGTATGATCGGTGCGCTAACCGGACCTGTAA 948
Db 1874 ACATACCGCGCGCGCCAGCTGACCTGTCTATCGGCGCCGCGNACAGGATGAGCGG 1933
QY 949 GCATTGGAACCCCTGACGTGTTTAAATATTCATCGGGAAGATCTTGTTATCAAGAGCGCT 1008
Db 1934 CGCTACCGCGACCCCGACCGGTTCGACATCTTCCGCGACGACCTCGCCACGACCTCGGCC 1993
QY 1009 TTTAGCGCGCGCGCGCGCATCTCGTTTCGGATCCGGCATTCATAACTGTCTAGAGCA 1068
Db 1994 TTTCTCGCGAGCTGCGCGCCACCTCGCCCTTCGCGCTCGCGCGCACCTTCGCTGGCGCG 2053
QY 1069 GCTTTTGC AAAAACAATAATGCAATTTGAGCTAATATTTGCTGCTGAATAAGATCGGAAT 1128
Db 2054 CTCCTGGCCACGCGGAGTTCGAGGTGGCTTGAACCACTTCTGACGCCATGCCGAC 2113
QY 1129 ATCAGATTAGAGGAAGATTTTGTATGCTGAGTCCGGTCTGTATACACCGCGA CTTGTT 1188
Db 2114 CTGCGCTCGCGACGCGGACGATCTCGTTGAACAGGGGCTCTTCAACCGCGCGCGAAG 2173
QY 1189 TCATTCTCTGT 1199
Db 2174 ACGTGCCTGT 2184

RESULT 7
SC0939106
LOCUS
DEFINITION
ACCESSION
AL313171 AL313424 AL3136149 AL1317165 AL445945 AL449216
AL591823 AL593822 AL645982
AL939106.1 GI:24413728
Streptomyces coelicolor A3(2)
Streptomyces coelicolor A3(2)
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1
Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitz,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,B., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorek,A., Woodward,J., Barrell,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
Nature 417 (6885), 141-147 (2002)
21996410
PUBMED
1200953
2 (bases 1 to 314100)
Bentley,S.D.
Direct Submission
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
On or before Oct 26, 2002 this sequence version replaced
GI:20520920, GI:20520898, GI:20520901, GI:20520903, GI:20520904,
GI:20520905, GI:20520906, GI:20520907, GI:20520926, GI:20520909,
GI:20520910, GI:20520805, GI:20520706, GI:20520789, GI:20520843.
Location/Qualifiers
1. 314100
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"

FEATURES
source
```

```
RBS
gene
CDS
misc_feature
misc_feature
gene
CDS
```

```
53. .56
67. .945
/gene="SCO0550"
/note="synonym: SCF11.30"
67. .945
/gene="SCO0550"
/note="SCF11.30, possible membrane protein, len: 292 aa;
similar to TR:Q9ZBQ4 (EMBL:AL034447) Streptomyces
coelicolor putative transmembrane protein, 396 aa; fasta
scores: opt: 677 z-score: 505.6 E(): 8.6e-21; 43.7%
identity in 309 aa overlap. Contains possible hydrophobic
membrane spanning region"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB59607.1"
/db_xref="GI:6137052"
/db_xref="GOA:O9RK60"
/db_xref="TrEMBL:O9RK60"
/translation="MGSRAARPSRHARPGGLSRGAALALGAVTLTAGVITLAV
GDGDDGASDRVTANAADOASNAEEAAGAAPLASASPSATAGSRSPSPAPKRTT
TKTASPTAPKASAPAKPVRTAKSGDRSGGSGSGSGSGSGSGSGSGSGSGS
SGSGTGAAGAAOVLALVNEERAAAGCSPVTANSRLTRAADDYSDVMASGVMGHTGPDG
STMSRVAAGYQWSALGENTIAQGQADAASVMEAWMNSPGHRENILNCSFKELGVSFH
PDGQPPWTQDFGTGR"
complement(1031. .2359)
/gene="SCO0551"
/note="synonym: SCF11.31c"
complement(1031. .2359)
/gene="SCO0551"
/note="SCF11.31c, possible histidine kinase protein, len:
442 aa; similar to TR:O50497 (EMBL:AL020958) Streptomyces
coelicolor two-component system sensor, 481 aa; fasta
scores: opt: 401 z-score: 451.3 E(): 9e-18; 30.2% identity
in 367 aa overlap. Contains Pfam matches to entries
PF00512 signal, Histidine kinase and PF00672 DUF5, domain
found in bacterial signal proteins. Contains possible
N-terminal signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative histidine kinase protein"
/protein_id="CAB59608.1"
/db_xref="GI:6137053"
/db_xref="GOA:O9RK59"
/db_xref="TrEMBL:O9RK59"
/translation="NRRALGIALATVSMVALSVLPLVLRQARDVTTAAEQRA
AALSPVALTTRADVQQAELGADOLVRLPDGFGVTGTPHAEALDRAVRGRET
LAVDTADGWAYLOPVLRGDRVAVVPAADLTRGVTAQSGVMALLAVGLVGSVL
VADRIGARVRSRGLKASLALSGDGLGVRVPPGPELQAGAAFTNADRVDDL
AVERELVADLSHRLTPTLALYLEADRMGATSAALRVTEAGQLRELDSTIAAKTPT
LAAAGLGTAGTAKPCDVAEYVAMRLDFWSVLAQAQDPFERSLTPRPAPVAFPEDD
LAAVVDALIGNVFRHTPQCTRFVVRVRRDRHRLTVDDAGPGVADPEALTRGVSVG
GSTGLGLDIVARAARTADGELTITRAPLGGARVTVSFALADARAAGDGR"
complement(1070. .1672)
/gene="SCO0551"
/note="Pfam match to entry PF00512 signal, Histidine
kinase, score 39.60, E-value 7.1e-08"
complement(1712. .1921)
/gene="SCO0551"
/note="Pfam match to entry PF00672 DUF5, Domain found in
bacterial signal proteins, score 36.90, E-value 4.5e-07"
complement(2356. .3072)
/gene="SCO0552"
/note="synonym: SCF11.32c"
complement(2356. .3072)
/gene="SCO0552"
/note="SCF11.32c, possible response regulator, len: 238
aa; similar to TR:CAB54568 (EMBL:AJ006392) Streptococcus
pneumoniae response regulator Rro2, 234 aa; fasta
scores: opt: 583 z-score: 693.9 E(): 2.8e-31; 43.2% identity in
227 aa overlap. Contains Pfam matches to entries PF00486
trans_reg_C, transcriptional regulatory protein and
PF00072 response_reg, response regulator receiver domain"
```


QY 1135 TTAGAGGAAGATTTTGTATGTCGATCCGCTCTGTATACACGGCGACCTGTTTCACATT 1194
 |||||
 Db 36314 ACCGAGGACGGCTTCGAGTGTGCGAGCGCGCTGTTTCAACCGCGCCGCGAGTCGCTG 36373
 |||||
 QY 1195 CTCGT 1199
 |||||
 Db 36374 CCGGT 36378

RESULT 8
 AEO12559
 LOCUS AEO12559 AE009442
 DEFINITION Xylella fastidiosa Temeculal, section 7 of 9 of the complete genome.
 AEO12559.1 GI:28057550
 SOURCE Xylella fastidiosa Temeculal
 ORGANISM Xylella fastidiosa Temeculal
 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xylella.

REFERENCE 1 (bases 1 to 30085)
 AUTHORS Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorri, H., Tsai, S.M., Carrer, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Blanco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Fenille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., Oliveira, A.R., Rosa Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P.
 TITLE Comparative Analyses of the Complete Genome Sequences of Pierce's Disease and Citrus Variegated Chlorosis Strains of Xylella fastidiosa
 J. Bacteriol. 185 (3): 1018-1026 (2003)

JOURNAL 12533478
 PUBMED
 REFERENCE 2 (bases 1 to 30085)
 AUTHORS Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorri, H., Tsai, S.M., Carrer, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Blanco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Fenille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., Oliveira, A.R., Rosa Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P.
 TITLE Direct Submission
 JOURNAL Submitted (17-DEC-2001) Instituto de Biociencias, Universidade de Sao Paulo, Rua do Matão, 277, Sao Paulo, SP 05508-900, Brazil
 FEATURES
 source
 1. 30085
 /organism="Xylella fastidiosa Temeculal"
 /mol_type="genomic DNA"
 /strain="Temeculal"
 /db_xref="taxon:183190"
 /note="Pierce's disease strain"
 87. 1022
 /gene="apbE"
 /locus_tag="PD1557"
 87. 1022
 /gene="apbE"
 /locus_tag="PD1557"
 /note="identified by sequence similarity; putative; ORF"

located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="thiamine biosynthesis lipoprotein ApbE
 precursor"
 /protein_id="AAO29399.1"
 /db_xref="GI:28057551"
 /translation="MGTTWSVKLAAPRRLDLHPLHNAIQARLDQIVAEWSTWEPNSHI
 SRFNAPVRSWHPLFNDPLVRLTALDIAMTSNGAFDPTVGLVWLGFGAYSGNHKQ
 IPTPEAITLATRVGQHLERTTDCGHMQPGTGLDLSAIGKGVSDVAFAATLHAKGV
 HBAVDIGGELYGCHKPDGTPWSLVEMDLTHOYNPLPPCIIQLDGLAVATSGDRW
 HFEHGORYTHITDPHQGTIPDPAALVMTIASSAMHADAWATALTILGREAGLALA
 NTLGLAVYLEHNDADFAYYSPAFTRLLHKPATY"
 complement (1752..2081)
 /gene="comE"
 /locus_tag="PD1558"
 complement (1752..2081)
 /gene="comE"
 /locus_tag="PD1558"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="DNA transport competence protein"
 /protein_id="AAO29400.1"
 /db_xref="GI:28057552"
 /translation="MKSFAMRLKSLLLFLVLAFTVHAANKVDINTASAEEMDKVLVNI
 GPSKASATKYREENGPFKSVBELALVKIGMKTVERNRLIEIGTRMAPAKHAKGTE
 LKSVGKR"
 2603..3625
 /locus_tag="PD1559"
 2603..3625
 /locus_tag="PD1559"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAO29401.1"
 /db_xref="GI:28057553"
 /translation="MTSSDFSDTTRATQRLQWIRTTADSTATLQASSDAGFRSVWR
 STSNGLSHLDNDAPHLNPNQWLRHALLTSGGVVRPHILAQDLQLEAGFLLEDGLIP
 NLNRTDLALQVQOQMDNALTPRVLVHRDFMRNLMLTTDGVTVLDFDCTVG
 PVADPVSLKDTSVSWPLRVDRWLTHYHARAANAAPVQTLPHFLRDADWMGQVRH
 LKNLGI FARLHYRDGKSWYLENIPRFISYLEILLPHRPTLAPLAEILIEHRIKPAALAA
 MITEST"
 3622..4344
 /locus_tag="PD1560"
 3622..4344
 /locus_tag="PD1560"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="nucleotidyl transferase"
 /protein_id="AAO29402.1"
 /db_xref="GI:28057554"
 /translation="MKALIFAAGIQRMRLNTYTPKLLCAGGEPILVWNLKLAAL
 GISEVVINTALWSEQFPBILGQGRFGLRFLYSNEGSLPLETGGGMLHALPILGNAPP
 LAINGDIWTDADLTLPTEPVGDHMLVNNPNPEYHPQGFVLQADSSVLDRTFPGIPTL
 TFGAGIYRSQLADWRNIIGDTPDTHAQPPRFKPLAPLLRAAMRSRGRIHTHRRQWT
 DVGTPQRLHALDTWLRSPEARF"
 complement (4527..5228)
 /locus_tag="PD1561"
 complement (4527..5228)
 /locus_tag="PD1561"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="replication related protein"
 /protein_id="AAO29403.1"

/db_xref="GI:28057555"
/translation="MSVSLPLALRYSSDQRETYLHAPSGLIAOLRALVDCSREWI
YLVGSGSKTHLALCAAQVQSVAYLQVAVGRLRLDALSGCVLVLDGL
QIAVGRREIALFHNRAAGITLLTYARAI PDGLGLTLPDURSRVAQCIRALP
TLDDVGRASVLRERARQRLMLDEAAIEWLLSRGRDLGSLVMLLERDLRESLATKR
VTIPFLRHVLAEGGG"
complement (5225..6382)
/gene="perM"
/locus_tag="PD1562"
complement (5225..6382)
/gene="perM"
/locus_tag="PD1562"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="permease"
/protein_id="AA029404.1"
/db_xref="GI:28057556"
/translation="MMFSPVEIARPLRLKXVVLVGGVLMWMLLTPVLIPVLSVF
LGNLADPLVRVWGVGRAMAVCFVFLVLLLVLMIFVPLIERQVTLINTFPQ
MHWVNTVPIWEGTQSLRQLTWLDPEQMIOMHSHNMEQAGVGARLFGVYSRGFV
MYTWLNLALFIFYLRDWDKWRVAAAIPRAYIGTISRLEANEVLGAPIRG
QFLVMLGAIYAVGLISVGLRLGLLIGLIAGLISFIPYLGAITGVWLALIAVIVQH
GLDOLLVLGVVFGVGLLESYVLTPIVIGDKILHPVAVIFSVMAGQLFGVGM
LALPVAVNVLLRYAHRYIESEFYKGYLAVALHECKSDENCDGKDAV"
complement (6379..7551)
/locus_tag="PD1563"
complement (6379..7551)
/locus_tag="PD1563"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA029405.1"
/db_xref="GI:28057557"
/translation="WAMPYFCQQLFCISHAMELSMRVRFELFALCAVLVPVAVA
QRPHIQSDVAKEQYIEAEVFNQSGSDSRKARALISVLGKTVGDRSNAKLRP
EVSQALRDAAVSDYDRQDSGSGGAPTSRTMLTVHFRDEYNALVGLALPLWP
QRPFPVLWLAIDDSGFPRLVGVQSRVSVLDRAIERGRGLGPAGRSERQALVTA
IWRDVSALVSATRYTSMQLVGLKSRGNGWAADWIFVGGEGVSWSSSNADALR
VNADGADGADALVKYAKVPLTGPVYRVGIRGSRADDYLRVSAALQRPVVRSM
IPKASAELELYLDLMTGITGLNMLGDSALQPVAAATLSPQSEGLVEYVLK"
7602..8657
/gene="purM"
/locus_tag="PD1564"
/note="synonym: purG"
7602..8657
/gene="purM"
/locus_tag="PD1564"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="5'-phosphoribosyl-5-aminoimidazole synthetase"
/protein_id="AA029406.1"
/db_xref="GI:28057558"
/translation="MTTPPPSTPLTYRDAGVIDAGNALVERIKPLVKRTFRPEVM
SGLGFGFALNLAGKPELVGTDGVTGKLKLAQLNRHTIIGIDLVAMCVNDVLV
QGAEPFLFDYFATKLDIDTATAVISGALCEOSGCALIGETAEMPDMPYPCGEYD
LAGFCVAVEKSQLDGSQVREDDVLIGIASGPHSNGYSLIRRIYERAGSPADLDIH
GRLDIDLTPALYKPLKULHSHDAIHAHAHITGGGLTENIRVIPPNLGURID
ANAWQPPVFOVQREGLADTEMWRTFNCIGFGVLVATPNQVAPLQALNDQGLAHV
QIGRVVTPVDNRVHIG"

Query Match 9.6%; Score 116.2; DB 1; Length 300885;
Best Local Similarity 49.4%; Pred.No. 6.8e-24; Mismatches 308; Indels 0; Gaps 0;
Matches 301; Conservative 0; Misamatches 308; Indels 0; Gaps 0;
375 TGATCTCGTCAATGATTTTGGAAAGACGTTTGGCGGTGCGTCACGATGACATGCTCGG 434
DB 156009 TGATCTTGTGCTGAATTTGCCTTTCCCTTCCATGCAGATCAATTTGCAAAATGATGGA 156068

QY 435 GCTGGATAAAGAGACCATGAAAAAATCTCTGAGTGGCACAGCGAGTTGCCGATTTTAT 494
|||||
DB 156069 TGTGATATTGAGATGCGGTCACCTCGGTATGCGGTGAGCAAAATTGCAGAGTACT 156128
QY 495 CACGAGTAGTCTCTCAATCTCTGGAAGCGCGGGGCACATTCGTTTATGGTGAGCAGCAACAGCT 554
DB 156129 CGACCTTCGCGCAATGTCGCGCGATGAATTAGTACACGAAGCACTGCTACGAAGAGCT 156188
QY 555 TTCCCAATATTGATGCGCGGTCAATTAAGAAGAGCTGCGGTCAATTCGCGGATCAGATTTAAT 614
DB 156189 TCGCAATATCTTTACGAAGTTGATTTAGCTACGCGCACCCATCTCTGGAACCTGACTTAAT 156248
QY 615 TTGGATCTCTATGCTTCTGATATGAAGGATGCGGTGTCGGACAAGGATATACTCGC 674
DB 156249 TTCCATGTTTCTGCGTCTGAGGAAGATGGCGAGAAACTAACCCATGATGAATCGTCTC 156308
QY 675 ACTGATCTTAAATGTGCTGTGTAGCGCAACGGAACCGGCTGATAAGAGCGCTGGCACTGAT 734
DB 156309 CAATGTGATTATGTTATTAAATCGCGGTTACGAACACCATCAATATGATTGGCAATGC 156368
QY 735 GATCTACCATTTGCTCAACAATCTCTGAGCAGATGAATGATGTTTGGTGGTGACCGTTCGTT 794
DB 156369 ATTGATTGCTTGCATCGCCATCCAGAGCAACTCGCACTGCTCAAGAGTGATCTGTCACT 156428
QY 795 AGTCCGAGAGCCATTGCGGAGACATTCGGTTTATAAACCOCGGTTCAGCTGATTCGCGG 854
DB 156429 GATGCCACAAGCGGTATCGGAATGCTTACGTTACACGGGTGCGGTACAGTTTCAAGTGG 156488
QY 855 GCAGTGTGCCAAGATACAGTGGTGGCGGTATGAAATCAAAAAGATACCAATGTTT 914
DB 156489 CGCGCGATGATGATAGAGTTGAAGCGGAGTTGTTCCCTCGCGCACAGTAGTATT 156548
QY 915 TTGATGATCGGTGCGGTAAACCGGACCTTGAAGCATTTGAACAGCTTGAGCGTGTTHAA 974
DB 156549 TTTGATGCTTGGTGTCTGCCAACCGTGATCCGCGACAATTTACTCATCTCCGACAGTGGA 156608
QY 975 TATTTCATCG 983
DB 156609 TATTACTCG 156617
RESULT 9
AE003889 10092 bp DNA linear BCT 04-JUN-2004
LOCUS Xylella fastidiosa 9a5c, section 35 of 229 of the complete genome.
DEFINITION AE003889 AE003849
ACCESSION AE003889.1 GI:9105205
VERSION
KEYWORDS
SOURCE Xylella fastidiosa 9a5c
ORGANISM Xylella fastidiosa 9a5c
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xylella.
REFERENCE 1 (bases 1 to 10092)
AUTHORS Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
Bueno,M.R., Camargo,A.A., Carraro,L.E., Carraro,D.M., Carrer,H.,
Colauto,N.B., Colombo,C., Costa,F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorfy,H.,
Faciniano,A.P., Ferreira,A.J., Ferreira,V.C., Ferto,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
and Marino,C.L.
TITLE The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis
NATURE 406 (6792), 151-157 (2000)
JOURNAL MEDLINE 20365717
PUBMED 10910347
REFERENCE 2 (bases 1 to 10092)
AUTHORS Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,

Alvarenga, R., Alves, L. M. C., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bove, J. M., Briones, M. R. S., Bueno, M. R. P., Camargo, A. A., Camargo, L. E. A., Carraro, D. M., Carter, H., Colaço, N. B., Colombo, C., Costa, F. F., Costa, M. C. R., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Pacinciani, A. P., Ferreira, A. J. S., Ferreira, V. C. A., Ferro, J. A., Fraga, J. S., Franca, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H. S., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J. P., Krieger, J. E., Kuramae, E. E., Laigret, F., Lambais, M. R., Leite, L. C. C., Lemos, E. G. M., Lemos, M. V. F., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, M. B. N., Madeira, H. M. F., Marino, C. L., Marques, M. V., Martins, E. A. L., Martins, E. M. F., Matsukuma, A. Y., Menck, C. F. M., Miracca, E. C., Miyaki, C. Y., Monteiro-Vitorello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L. O., Netto, L. E. S., Nhani Jr., A., Nobrega, P. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A. G., Pereira Jr., H. A., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M. Rosa, A. J., de Rosa Jr., V. E., de Sa, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C. R., da Silva, F. R., da Silva, A. M., Silva Jr., W. A., da Silveira, J. F., Silvestri, M. L. Z., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tshako, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J. and Setubal, J. C.

Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

Location/Qualifiers
1. .10092
/organism="Xylella fastidiosa 9a5c"
/mol_type="genomic DNA"
/db_xref="taxon:160492"
/clone="9a5c"
715. .1482
/locus_tag="XF0374"
715. .1482
/locus_tag="XF0374"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83184.1"
/db_xref="GI:9105206"
/translation="MTYNNISILSRINQACIQARKEFLFHLWPLVGLVMSFGLMG
LAGDOWADLLRLSGQWLLKQHWITEYLHVRGKVLISVAGLGLGLMLVRCRGL
AGCYRLASLYVVVTLVLTSLISLVKLTGMDPCWDLVRYGGPPFGLFESRGLK
ASGCFPAGHASAGYAVCLYFAAWGVCFAWRWGLMIGLVAGLIFGISQQLRGAHFLS
HDLSLTICWLVAGLFFYLFVSPTRVVSQAQRANT"
1479. .3137
/locus_tag="XF0375"
1479. .3137
/locus_tag="XF0375"
/note="similar to SP|30845 (percent identity: 36 %/query alignment coverage: 98.0 %/subject alignment coverage: 98.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="inner membrane protein"
/protein_id="AAF83185.1"
/db_xref="GI:9105207"
/translation="MSFSLVILVLRMRSEFNRVRPEVSTSVLGTSLFALACNT
MFWRSAMTSAGSIGFVSLALLVTYHALLGLVWVRSAKPLTLFVITAFATHY
MNSYVLDDLRNVFNDHRESRELTISALPLLFYAGVPIAVLWLRFRQRPWS
RALGRMLPLLVVVGAGGCMFSLKMLNDRVRLATPINYIMALKMSLND
SVLKSPKIPGEDAVATPRIIPSSRPLVIVVGTARQNLGNGVROTPPOLAON
DVINPDMHSGTNTVEVPCMFSPYGRNRYDESKIRHQSLHLVLERISTLRDN
QSGCKGVCGLSLQDDAKDPTLCYSGRCMDLILKDFVQSRKGLGDRVVVHQL
GSHGSPFYQRYFAVRFQSPCTCNLGSCEQVVAAYDNLSTLYDFLVTIGKLR
DMSDYDTAMTILYSDHGESLGEGLYHGMFYAIPAEOQTRVPMVWFQSKQFVQSRQID

gene
CDS
LSCVHQARQYADHNDLPSSVLGMQVKTALYERPHDLFATCEK"
complement (3449. .3610)
/locus_tag="XF0376"
complement (3449. .3610)
/locus_tag="XF0376"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF83186.1"
/db_xref="GI:9105208"
/translation="MRLSRITDFSLLKLRIRLRLQRIQAHYGGTLLDLVLITPEP
LAEAYENL"
complement (3751. .4959)
/locus_tag="XF0377"
complement (3751. .4959)
/locus_tag="XF0377"
/note="similar to SP|53554 (percent identity: 37 %/query alignment coverage: 94.3 %/subject alignment coverage: 95.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: -108"
/codon_start=1
/transl_table=11
/product="cytochrome P450-like enzyme"
/protein_id="AAF83187.1"
/db_xref="GI:9105209"
/translation="MKLTDLNPNPFLYPLYLRAQAPFVSGPNALMTGRYSVL
DSLHNRNMGKYMESBVRVYCDSDAEMFLQAFSRMITINPPATHRLGLVMQAFV
GRESMRPLADTAHQILDNEFEQKPSVDLVAEPFPMQLICKMDVDIGDVTIG
IAVSKIAKVPFSPMSADELHASTAYBELAQYFKLIELRKHGTDLSIFLRABE
DCKLTHDEIVSNVIMLIAGYETTSNMGALIHRRHLPOLAKDLSLMPQAVS
BCLRYDGSVQFTMRAMDDIEVEGELVPRGTVFLMGAANDPQAQTHPDQDITRK
CQRLOSFGAGTHCGYRLALIELECAITLFLERPLRLAHLDALNWNQSRNLGRVN
TLIVDLHAKN"
complement (5841. .6545)
/locus_tag="XF0378"
complement (5841. .6545)
/locus_tag="XF0378"
/note="similar to GI|5932372 (percent identity: 41 %/query alignment coverage: 80.8 %/subject alignment coverage: 79.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="thiamin-phosphate pyrophosphorylase"
/protein_id="AAF83188.1"
/db_xref="GI:9105210"
/translation="MEVSRIGSFQBIHDKTGTIVTPHPLLRPLMSQPRGIYLTPE
TDTARLIARTPLNGIYVLRNKLANTALRTEQAALLACROTGPLLLINDLLEL
AQTIQADGVHLGHMDSNPISARAQIGPHAIIGVSCNQTAKQAKAGASVGFQTF
YPSHTKTPYRATPELLRQTTHLGVPRVAIGLTPKNTAPIEAGALLAVISGIYSA
KNVTALAKYQSQFNI"
6547. .6720
/locus_tag="XF0379"
6547. .6720
/locus_tag="XF0379"
/note="similar to SP|242453 (percent identity: 54 %/query alignment coverage: 93.0 %/subject alignment coverage: 98.1 %); identified by sequence similarity; putative; ORF located manually"
/codon_start=1
/transl_table=11
/product="rubredoxin"
/protein_id="AAF83189.1"
/db_xref="GI:9105211"
/translation="MMWFTWCTCGFLYRBEFGPFBEGIAAGTWNEDVSEWVCPDC
GIGKADFFEMVITVG"
complement (6854. .7021)
/locus_tag="XF0380"
complement (6854. .7021)
/locus_tag="XF0380"
/note="hypothetical protein; identified by sequence

similarity; putative; ORF located using Glimmer/RBSfinder	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="AAF83190.1"	
/db_xref="GI:9105212"	
/translation="MLLKLNHVDFDKQSKLMTQKPSVHPKKNKSQKEIKLSHQKRD NNPRYKIFTYI"	
gene	7207.. 9792
/locus_tag="XF0381"	
CDS	7207..9792
/locus_tag="XF0381"	
/note="similar to Sp P03815 (percent identity: 69 %/query alignment coverage: 99.5 %/subject alignment coverage: 100.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"	
/codon_start=1	
/transl_table=11	
/product="ATP-dependent Clp protease subunit"	
/protein_id="AAF83191.1"	
/db_xref="GI:9105213"	
/translation="MRMDKLSRFQNALADAQSLAVGRDHTIIEPVHVSALLDQGG STPLMQAGNVPLRLERITEILALPKVSGQTVNVSFNSLSRLFHRDKLAQOHG DQFMASEVFLAVDDSGGLRAAGAEKKIEAIDKLRGGTIVOTNAEORQA LEKYITDLTARAESGKLDVPVIGRDEIRRTIQVLRRTKNNPVLIGSPGVGKTAIVEG	
Query Match	9.3%; Score 113.6; DB 1; Length 10092;
Best Local Similarity	47.5%; Pred. No. 2e-23;
Matches 338; Conservative 0; Mismatches 374; Indels 0; Gaps 0;	
QY	272 TTGTAGTGAAGAGCTTTATCGGTGACGCACTCGGATCATCTGTCTCCAATTGATTAAACAAA 331
DB	4673 TGGTAATGAAGCGTTTCACCGGCGAGAGAGTGAATCAATGCGTCCCTTGGCATCGATA 4614
QY	332 ATGCAGAAAACCTGTTAGCGCCTTATCTTGAAGAGGGAAAAGTAGTCTCGTCAATGATT 391
DB	4613 CGGCGCACCAATTGATCGCAACTTTGAAACAGAAACCATCCGTTTGATCTTTGTTGCTGAAT 4554
QY	392 TTGGAAGACGTTTGGCGTGTGCGTACAGATGGACATGCTCGGCTCGATAAAGAGACCC 451
DB	4553 TTGCGTTTCCCTTTCCCATGCAGATCATTTTGCAAAATGATGGATGTGGATATTGGAGATG 4494
QY	452 ATGAAAAAATCTCTGAGTGGCACAGCGGAGTTGCGGATTTTATCACGAGTATCTCTCAAT 511
DB	4493 CCGTCAACCTGGGTATAGCGGTGAGCAAAATTCGCAAGTATTCGACCTTCGCCAATGT 4434
QY	512 CTCCTGAAGCGCGGGCACATTGTTGTTATGTTGTCAGCGAACACAGCTTTCCCAATACCTTGATGC 571
DB	4433 CGGCCGATGAATTAGTACACGCAAGCACTGCTTACGAAGAGCTTGGCAATACTTTTACGA 4374
QY	572 CGGTCAATTAAGAACGTCGCGTCAATCCGGGATCAGATTTAATTTTCGATCCTATGTACTT 631
DB	4373 AGTTGATTGAGCTACGCGCGCACCCATCTTGGAACTGACTTAATTTTCTATGTTTCTCGCGT 4314
QY	632 CTGAATATGAAGCATGCGCGTGTGCGACAAAGGATATATCTCGCACTGATTTCTTAATGTGC 691
DB	4313 CTGAGAAGACGCGGAGAACTAACCCATGATGAATCGTCTCCAATGATTTAGTTAT 4254
QY	692 TGTTAGCCGCAACGGAACCGGCTGATTAAGACGCTGGCACTGATGATCTACATTTGCTCA 751
DB	4253 TAAATCGCGGTTACGAAACCAATCCAATATGATTGGCAATGCAATGATTGCGTTGCATC 4194
QY	752 ACATCTCAGCAGATGAATGATGTTTTTGGCTGACCGTTTCGTTAGTTTCCGAGAGCCATTG 811
DB	4193 GCCATCCAGAGCAACTCGCACTCCTCAAGAGTGAATCTGTCACTGATGCCCAAGCGGTAT 4134
QY	812 CGGAGACATTGGCGTTATAAACCGCGCGTTCAGCTGATTCGCGGCGAGCTGTGCCAAGATA 871
DB	4133 CGGNAATGCTTACGTTACGACGGGTTCGATACAGTTTACAGTGCAGCGCGCGGATGATGATA 4074
QY	872 CAGTGGTCGCGGTTATGGAATCAAAAAAGATACGATGTTTTTTTGTATGATCGGTGCGG 931
DB	4073 TAGAGGTTGAAGCGGAGTTGGTCCCTCGTGGCACAGTAGTATTTTGTGATGCTTGGTGCCTG 4014

QY	932 CTAACCGGACCTGAAGCATTGTAACAGCCTGACGCTGTTAATATTATTCATCG	983	
DB	4013 CAACCGTGATCCGCGCAATTCTACTATCCGACCACTGGATATTACTCG	3962	
RESULT 10			
LOCUS	AE017008	30090 bp	DNA linear BCT 16-MAY-2003
DEFINITION	Bacillus cereus ATCC 14579 section 11 of 18 of the complete genome.		
ACCESSION	AE017008	AE016877	
VERSION	AE017008.1	GI:29896738	
KEYWORDS	Bacillus cereus ATCC 14579		
SOURCE	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.		
ORGANISM	1 (bases 1 to 300090)		
REFERENCE	Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B., Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A., Chu,L., Mazur,M., Goltzman,E., Larsen,N., D'Souza,M., Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M., Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.		
TITLE	Genome sequence of Bacillus cereus and comparative analysis with Bacillus anthracis		
JOURNAL	Nature 423 (6935), 87-91 (2003)		
MEDLINE	22608415		
PUBMED	12721630		
REFERENCE	2 (bases 1 to 300090)		
AUTHORS	Candelon,B., Gailloux,K., Ehrlich,D.S. and Sorokin,A.		
TITLE	The number of ribosomal RNA operons in Bacillus cereus		
JOURNAL	Unpublished		
AUTHORS	3 (bases 1 to 300090)		
	Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B., Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A., Chu,L., Mazur,M., Goltzman,E., Larsen,N., D'Souza,M., Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M., Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-MAR-2003) INRA, Genetique Microbienne, Domaine de Vilvert, Jouy en Josas 78352, France		
FEATURES	1..300090		
source	/organism="Bacillus cereus ATCC 14579"		
	/mol_type="genomic DNA"		
	/strain="ATCC 14579"		
	/db_xref="ATCC:14579"		
	/db_xref="taxon:226900"		
	complement(109..660)		
gene	/locus_tag="BC3070"		
	complement(109..660)		
CDS	/locus_tag="BC3070"		
	/EC_number="3.4.21.89"		
	/codon_start=1		
	/transl_table=11		
	/product="signal peptidase I"		
	/protein_id="AAP10017.1"		
	/db_xref="GI:29896739"		
	LKNRVLVNKIGYSISGLDRFDIIVPHKGYDLVKRVIGLPGDVEYKNDLVYNG KAMEEVLKQFKEKAVGRVLTPTLEQITGKTVPQGVFLGDNREVSKDGRMFG ISEIEIVKGQAVFWPLEQVRL"		
	complement(684..1361)		
gene	/locus_tag="BC3071"		
	complement(684..1361)		
CDS	/locus_tag="BC3071"		
	/codon_start=1		
	/transl_table=11		
	/product="Copper homeostasis protein cutC"		
	/protein_id="AAP10018.1"		
	/db_xref="GI:29896740"		
	/translation="MLEVIATCLEADVKEIESAGGKRIELISSYTEGGLTTPSYAFIKKA VEVWSIPTHVIMIRPHAKSFTVTEESIEEMKEDIIVAQKLGAAGVVLGVLNERNEVDEE		

KLADLLSVVDGIVNTVTHRAIDDIKNPVEAMSTLKKPKFKVTHVLTSGGQGNIVENIPVL
TMRQVSEGOIQLVAGAGVTKENIKRLDETGITQAHVGTAVREGKSCFSEIDPNLVQ
ELVEI1K"
complement (1560..2267)
/locus_tag="BC3072"
complement (1560..2267)
/locus_tag="BC3072"
/codon_start=1
/transl_table=11
/product="Transcriptional regulator, Arsr family"
/protein_id="AAP10019.1"
/db_xref="GI:29896741"
/translation="NMVYNISYIAKLIAPETRAIILDCLMNQALPASELAYMAKVS
HPTISHSKLVGELLTVQGRHRYHLANOEAELVLEKIGTIAPTQVVRSLKQSS
QKQIRYARTCDYHLAGVIGTEIKLDRQFIILEDGYIIVTEQGWFLNFGINVD
EADTKRVPKFLDMSERYHISGWLGSIAKLLFEQGWITKTDKRAVHLTKGKVK
LLEDQLGIDMKTEERKA"
2337..2807
/locus_tag="BC3073"
2337..2807
/locus_tag="BC3073"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP10020.1"
/db_xref="GI:29896742"
/translation="WNAIDLSILNLKETRRSEKLNWLPDNLNWKDPPEAMSRGEM
IRHWSSTFYHMIKNNGSINDIRPYDDEPITCVKKEIELAQSHFTDFIGVQSIS
IAELDSRLDRSDVGQRYLGDMLLRIYHDAVHAGQFLQYLRMYNLERPLWD"
complement (3023..4172)
/locus_tag="BC3074"
complement (3023..4172)
/locus_tag="BC3074"
/note="similar to Acyl-CoA dehydrogenase, short-chain
specific; EC number 1.3.99.2"
complement (4327..4959)
/locus_tag="BC3075"
complement (4327..4959)
/locus_tag="BC3075"
/codon_start=1
/transl_table=11
/product="EpiH/GdmH-related protein"
/protein_id="AAP10021.1"
/db_xref="GI:29896743"
/translation="WKENKSKKRIRIFOVFLMIGSAILVSYAAVDIWSYRFXANDD
VTDAGIVLGAASWNGKSPVFKERINHAISLYKNGNIKKIIFTGTGTFEALEART
ARYAMKQGVKEDILIEKSLFTENLKNKQVGIENGIOITYTIVSDPLMKRAMRI
AKHIMEAYSPTPSAYKTLDTETPFPPKELFSYIGYVTSIPLKALGKD"
complement (5289..5876)
/locus_tag="BC3076"
complement (5289..5876)
/locus_tag="BC3076"
/EC_number="2.3.1.-"
/codon_start=1
/transl_table=11
/product="Acetyltransferase"
/protein_id="AAP10022.1"
/db_xref="GI:29896744"
/translation="WEIHERAKRLMDNTVETLPSIVEGRNRIWAYILSKQDSVQD
MOQYQKALFGRGQTPIPIVDQSNKIYVSTRLYNISVEDKTVELGQTVHPSPVQ
RTSINTEYMLQVAFELKHLRLVQIKTDARNEKAQRAIERLGAVKEGVLRNERKL
NGYVRDVAVVSIISSSEWSPVSKLEKLELYEKL"
complement (6268..7596)
/locus_tag="BC3077"
complement (6268..7596)
/locus_tag="BC3077"
/codon_start=1
/transl_table=11
/product="hypothetical Membrane Spanning Protein"
/protein_id="AAP10023.1"
/db_xref="GI:29896745"
/translation="MSEIELKSNVINTKREKGIILRFILASLVGVMPFVFPVINGAS
SIMIDHIVSWIRASVPGVVPYVYALFVMAIGAIYPFYTKKNASIVDICFSILKVGW

gene
CDS

gene
CDS

gene
misc_feature

gene
CDS

gene
CDS

gene
CDS

gene
CDS

FGILYCLKVPAWFPADIGPFLYEKLVISVSLVPIGSALFALLVGVGLLEFIFGTFCC
RPIMPLNWTGERSAIDAVASVYSYSLALLITNRVYKEGKTYTTEAAIIATGFTSTVS
ATFMIIDITLIDIMHNMVYFWTTLVTFIVTAITVRIPPLSRKPDYVTEGFPPEPV
YKEKLERAWEDALEVSKSAPSIMKNVIANVLKGTIMTMGLIPSMVSAGLIGVILAKF
TPIPDWISYIFYPWTLLQLEADLAARASVGIEMELPSLLVMSAPLVTKPVIAVV
SVSSILPFSASIPCLISLDIPKSELILYVQRTILLIITPAYALLL"
complement (7708..8586)
/locus_tag="BC3078"
complement (7708..8586)
/locus_tag="BC3078"
/EC_number="2.3.1.-"
/codon_start=1
/transl_table=11
/product="Acetyltransferase"
/protein_id="AAP10024.1"
/db_xref="GI:29896746"
/translation="MHPINVSLEKLIQORPEWAHLEVKPKVFGYDNRTFHLGDEM
SVRLPSDVAYAPOVEKESWLPILNKGSLSPSTPLAKGNPSEAPLPSINKMLGE
TVTKQNVDRDLNEPADLGSFLVELQSINASNGPLAGTHNFRYGGGLISYVDEARVAIE
NNKQVDEALLKHLNVALSSTWDRKPVWVGDVAPGNLLVKDGKLCVDFGILGVG
DPACDANAWTFFDENSRNVFKEVLRMDSETWNRARGWALWKALITYDANRESNEKVA
EESYRFIQIVDDYER"
complement (8587..9102)
/locus_tag="BC3079"
complement (8587..9102)
/locus_tag="BC3079"
/EC_number="2.3.1.-"
/codon_start=1
/transl_table=11
/product="Acetyltransferase"
/protein_id="AAP10025.1"
/db_xref="GI:29896747"
/translation="MERNISKEYAIRIATENESDSIIILLKEVAQMLQYKEVDQWQYL
LGEHTABILECIREKTYTVLKEAEIVGTVTVPKQNEWDEHIFCKEVSNSLYIHR
FAVKRYKYGNGIGEWILRWVENVQCDKEYLKLDCVGHNRITLNDPVKKNSEYVVGSTD
GLSKPKKRRM"
complement (9126..9590)
/locus_tag="BC3080"
complement (9126..9590)
/locus_tag="BC3080"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP10026.1"
/db_xref="GI:29896748"
/translation="MKRGLTVDLGDYDNLMLKVLKPIDITVDFMRVRDEESYLIARG
ELDEALPFEESPVVEGLAKKLKVDNIYLI PADLKAVPKGEIADIDETVEEFKESKC
EVVILVAQAQYIYKADQKSELHNAVNOGTYVEYITDENDGRTRLSVW"
complement (9807..10244)
/locus_tag="BC3081"
complement (9807..10244)
/locus_tag="BC3081"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP10027.1"
/db_xref="GI:29896749"
/translation="MMVNSNYDWRONDKLSIDIEKAINGEYSAISCVAKLANMAPN
QVEQKQLEIRNDEIKHFNFIQIYNTLTGRQPKQITDCDCEYTLQGLEFALQDEQK
TVDFYLEISDSTSDVHLKELLRIADAQNHAWFLLYTFVASK"
10551..11057
/locus_tag="BC3082"
/locus_tag="BC3082"

gene
CDS

gene
CDS

gene

Query Match 7.5%; Score 90.8; DB 1; Length 300090;
Best/Local/Similarity 64.0%; Pred. No. 6.5e-16;
Matches 137; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 916 TGTATGATCGGTGCGGCTAACCGGACCTTGAGCATTTTGAACAGCTGACCTGTTTAAT 975
|||||
Db 59764 TGTATGATGGTGGCAACAATTTGTGACTTAACGATTTTGAAGGCTGATAAATTTCAAC 59705
|||||
QY 976 ATTCAATCGGGAAGATCTTGGTATCAAGCGCTTTTAGCGGCGCCGCCCGGATCTCGCT 1035
|||||

Db 59704 GTTATCGTCCTGACATTCATATAAAAAGGCATTTAGTGGTACTGCAAGACATCTTGCT 59645

QY 1036 TTTCGATCGGATTCATCACTGTGTAGAGAGAGCTTTTCCAAAACGAATCGAATTT 1095

Db 59644 TTTCGATTAAGTATCTATTAATTGTGTAGAGTAGCATTTTGCAAAATTAAATAATGAAATA 59585

QY 1096 GTAGCTAAATATTGTGCTGCGTAAGATGCGGAATA 1129

Db 59584 GATTCACCATTAAGATATATTAGCAGGAAAA 59551

RESULT 11

AE003887

LOCUS Xylella fastidiosa 955c, section 33 of 229 of the complete genome. BCT 04-JUN-2004

DEFINITION Xylella fastidiosa 955c

ACCESSION AE003887

VERSION AE003887.1 GI:9105175

KEYWORDS

SOURCE Xylella fastidiosa 955c

ORGANISM Xylella fastidiosa 955c

REFERENCE 1 (bases 1 to 10261)

AUTHORS Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frolman,L.R., Garner,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Ho,P.L., Hohnsels,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., and Marino,C.L.

TITLE The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis

Nature 406 (6792), 151-157 (2000)

JOURNAL MEDLINE

PUBMED 20385717

REFERENCE 2 (bases 1 to 10261)

AUTHORS Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., Bueno,M.R.F., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frolme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hohnsels,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuranae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.C., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Silva Sasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silva,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.

Direct Submission

Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

Location/Qualifiers

1. .10261

TITLE

JOURNAL

FEATURES

source

/organism="Xylella fastidiosa 955c"

/mol_type="genomic DNA"

/db_xref="taxon:160492"

/clone="955c"

/complement(189..341)

/locus_tag="XF0348"

/complement(189..341)

/locus_tag="XF0348"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF83158.1"

/db_xref="GI:9105176"

/translation="MEIVDAMRSHTLNDVTFPPSPNCNGSATYWWERSQITRCALQD FYISPV"

384..557

/locus_tag="XF0349"

384..557

/locus_tag="XF0349"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF83159.1"

/db_xref="GI:9105177"

/translation="MLALLTCTCAGMKMQLYNYKSPLIATNALPHQLSAHTVVIYRC LMRPLGNSMGFW"

564..659

/locus_tag="XF0350"

564..659

/locus_tag="XF0350"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF83160.1"

/db_xref="GI:9105178"

/translation="MPGLHLDFFHLNRDPYRNAGFMIDASVOLH"

/complement(707..931)

/locus_tag="XF0351"

/complement(707..931)

/locus_tag="XF0351"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAF83161.1"

/db_xref="GI:9105179"

/translation="MYASHLQWHRSLRIGSKGIRTSDLNLYQSSCVESLKRTHYPQA AIFCNQWIAVLFTPDHPMVLNRNRPGRQ"

1171..3378

/locus_tag="XF0352"

1171..3378

/locus_tag="XF0352"

/note="similar to SP|P17580 (percent identity: 47 %/query alignment coverage: 93.7 %/subject alignment coverage: 98.1 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="pentaphosphate

/protein_id="AAF83162.1"

/db_xref="GI:9105180"

/translation="MMLLIGHWSLPNMTWVSLIQTDHAALPLADAKVPDYFIQLERMA AHLTHEQMLLRRAWMMGAAHAGTKRSGSEFYITHPVAVACVLADLGMVDVETLAAI LHDITIEDTSVTRTEILAAEFGTVADLVGVTKLKFKRNQEAASAEFRKMLLAISR DLRVIMIKLADRLHNMRTLGAQSIEARSRIARETLEIYAPIAQRIGLMAIMKSELQNLG

gene	F0ALHPWHAIIKEHILNQPVVREAREMVELQSLRQLQEGVQHRVSVRIKTPWSI NKHNDNSFDQWMDVFGFRLVKSVDYCHALGIVHGTLPDLGRFDFPFIAPKTING YQSLTFLVFGPSIEVQIRTEEMDMIAERGAIAHWAYKCGSDSPNQAQRVHAWIA ELLESORTAISMEFLDNFQVDLPFDEVLYFTPKGI FALPKNSALTDFAVAVHDLQIG NHAVSRVDKSLMLRKLSSQGVTEIITARSALPKPOMLEFVSVSSKARTAIRHDLQIG LOHEDAVOLGHRMLDRALEMDSSLERLPKGRDLAFLEHHPQRLEALLADIALGNWM PNQTAQALITYTELLNGTPIESHKBLILINGSERGVISFANCCQPIPGDDVMVGHYITG KGIHVHRLDCPNLNRKSPERWVPIDWDLNVNSGDYDASLLVEAENRTGVLAQLAALAI AQSHSNIHVDYLERDFNAALLRFSIQVRDRHLAEVMMRLRLPAVQSVYRQ" 3592..3975 /locus_tag="XF0353" /locus_tag="XF0353" /locus_tag="XF0353" /locus_tag="XF0353" /note="similar to SP P40431 (percent identity: 54 %/query alignment coverage: 99.2 %/subject alignment coverage: 99.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: 180" /codon_start=1 /transl_table=11 /product="translation initiation inhibitor" /protein_id="AAF83163.1" /db_xref="GI:9105181"		CDS	VGDIAVQARRARDNLRARAVRANGSLKIVRLGLYLDLEQPAVVNVMQEFQAPFP AKSTIQVGLPKGADFVDVAVNVID" 3983..6139 /locus_tag="XF0354" /locus_tag="XF0354" /locus_tag="XF0354" /locus_tag="XF0354" /note="similar to SP P24230 (percent identity: 55 %/query alignment coverage: 97.8 %/subject alignment coverage: 101.3 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="ATP-dependent DNA helicase" /protein_id="AAF83164.1" /db_xref="GI:9105182"	
	VCDLAVQARRARDNLRARAVRANGSLKIVRLGLYLDLEQPAVVNVMQEFQAPFP AKSTIQVGLPKGADFVDVAVNVID" 3983..6139 /locus_tag="XF0354" /locus_tag="XF0354" /locus_tag="XF0354" /locus_tag="XF0354" /note="similar to SP P24230 (percent identity: 55 %/query alignment coverage: 97.8 %/subject alignment coverage: 101.3 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="ATP-dependent DNA helicase" /protein_id="AAF83164.1" /db_xref="GI:9105182"			WLLPLRYEDRRLTPIAELQDQVAVQIEARVEASPRYRPLRLVLTDDAHILT MRPFPRQAIFVTVGTVRAYGVKLGQGWELVHPSYRILAPGEDVRLNDCIDPV YPIEGVGPAIVQLIREALEHLPDVALELLPDWLSDLGLPMLRTALLIMHRPPLN TDIARLMVTHPAQSLLEELAHQLSLRQRIALQHSAPALPDGAALVASLLHAL PPLHAAQQVFAQIAADYAVKVPMLRVQGVSGKTVAALAAALLAVEQNKQVALLA APTALLAEHDLNLDWLEPLGVVHVAALKVTKARQLQALNDVASDAQVVGTHAL MOESVFDHLALTIYDEQHRFGVHORLLLRDKGATAGIVPHQLVMTATPIPTLMAAA YADLVSVIDENPGRTPQTVNLSAERPPELVERILACAGROVYVWCTLLIDESOT EAPQPPSNDIEHRSEVQAQATYEAUSQLPQVGVGLVHGRKAAEKQRTNRAPKC NEIDLVAITVIEGVDPVNSALIMIENAEFRGLAQHQLGRVGRGVSVSSCVLLYQ PPLSLARQLRLEIMQTDGFAIAEKDLRLGPGELLGTQTLGLAAFRVADILDAHL LPRVYSLANMLLDESPLAQADVIARWIGSAVYASA" 6214..6360 /locus_tag="XF0355" /locus_tag="XF0355" /locus_tag="XF0355" /locus_tag="XF0355" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF83165.1" /db_xref="GI:9105183"	
gene	VLRN" 6723..7925 /locus_tag="XF0356" 6723..7925		CDS	VLRN" 6723..7925 /locus_tag="XF0356" 6723..7925	
	Query Match 7.4%; Score 89.4; DB 1; Length 10261; Best Local Similarity 46.4%; Pred. No. 8.1e-16; Matches 291; Conservative 0; Mismatches 336; Indels 0; Gaps 0; 376 GATCTCGTCAATGATTTCGAAAGACGCTTTTCGGGTGTGGTGCATCGATGCATGCTCGGG 435			376 GATCTCGTCAATGATTTCGAAAGACGCTTTTCGGGTGTGGTGCATCGATGCATGCTCGGG 435	

Db	7113	GATCTTGTACTGAATTTCTCTTTCCCTTCCCATGCGGATCATTTTGCAGGATGATGAT	7172
Qy	436	CTGGATAAAGAGACCATGAAAAATCTCTGAGTGGCACACGGAGTTGCCGATTTTATC	495
Db	7173	GTGGATATCAGCGACGCTATCAGTCTAAGTGTGGCGGTGAGCAACATTGCCAAGGTATTG	7232
Qy	496	ACGAGTATCTCTCAATCTCTTGAAGCGGGGACATTTGTTATGTGTGACGGAACAGCTT	555
Db	7233	GACCTCGCGCGATGTCGCCCGATGAATTGGTGTGACGCAAGTGTGCTTACGAAGAATA	7292
Qy	556	TCCCAATACTTGAATCGCGGTGATTAAGAAAGCTCGCGTCAATCCGGGATCAGATTAAAT	615
Db	7293	GGCATTACTTCAAGGTGTGATCGAGTGCGCCGCGGACGCTGGAACTGACTTGAAT	7352
Qy	616	TCGATCTTATGTACTTCTGAATATGAAGGCATGGCGCTGTTCGGAACAAGGATATACTCGCA	675
Db	7353	TCCATGTTGTCGCGTCTGAGGAGGAGGGCCAGAACTTGACGCATGATGAAATCGTGTCC	7412
Qy	676	CTGATCTTAAATGTCTGTTAGCCGCAACGGAAACCGGCTGATTAAGACGCTGGCACTGATG	735
Db	7413	AATGTGATTTTGTGTTTATTTGGTGGCTACGAGACCATCAACATGATTGGCAATGCA	7472
Qy	736	ATCTACCATTTGCTCAACAATCTCTGAGCAGATGAATGATGTTTGGCTGACCGTTGTTA	795
Db	7473	TTGATGCGTTCATCGTCAATCCGAAGCAGCTTTCGACGACTCAAGAGTGAATTTGTCGCTG	7532
Qy	796	GTTCCGAGAGCATTGCGGAGACATTTGGTATTAAACCGCGGTTCACTGATTTCCGCGG	855
Db	7533	ATGCCGCAAGCATATTGGAATGCTTACGTTACGATGATCGTACAGTTCACGATGCGC	7592
Qy	856	CAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATAAATAAAGATACGATTTGTTTT	915
Db	7593	GCGGCGATGATGATGTGAGCATTTGAAGTGTATGTAGTACCTCGTGACGAGTTGTTTT	7652
Qy	916	TGTATGATCGTGGCGCTAACCGGACCCCTGAAGCATTTGAAACGCTGACGCTGTTAAT	975
Db	7653	TTGATGCTTGGTGTCTTAACCGTGTATCGCGGCAATCACTGATCCGATCATCTGGAC	7712
Qy	976	ATTATCATCGGAAGATCTTGGTATCAAG	1002
Db	7713	ATCACCGGAAGCAAGGACGCTCCAG	7739
RESULT 12			
AL646080			
LOCUS	190050 bp DNA linear BCT 11-APR-2003		
DEFINITION	Ralstonia solanacearum GM11000 megaplasmid, complete sequence; segment 5/11.		
ACCESSION	AL646080 AL646053		
VERSION	AL646080.1 GI:17431113		
KEYWORDS			
SOURCE	Ralstonia solanacearum		
ORGANISM	Ralstonia solanacearum		
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Ralstonia;		
AUTHORS	1 Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J. C., Cattolico, L., Chandler, M., Choise, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, F., Siguiet, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. and Boucher, C.A.		
TITLE	Genome sequence of the plant pathogen Ralstonia solanacearum		
JOURNAL	Nature 415 (6871), 497-502 (2002)		
MEDLINE	21681879		
PUBMED	11823852		
REFERENCE	2 (bases 1 to 190050)		
AUTHORS	Boucher, C.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston		
	Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie		
	Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,		

BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CRPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Bouchet@toulouse.inra.fr
<http://sequence.toulouse.inra.fr/R.solanacearum.html>.

COMMENT

FEATURES

source

1..190050

/organism="Ralstonia solanacearum"

/mol_type="Genomic DNA"

/strain="GMI1000"

/db_xref="taxon:305"

/plasmid="megaplaemid"

51..362

/gene="RSp0643"

/note="synonym: RS05579"

51..362

/gene="RSp0643"

/function="miscellaneous; unknown"

/note="Product confidence : hypothetical

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Framed

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="HYPOTHEITICAL PROTEIN"

/protein_id="CAD17794.1"

/db_xref="GI:17431114"

/db_xref="GOA:Q8XS38"

/db_xref="TrEMBL:Q8XS38"

/translation="MKTSPNPAAAGPASDPEAAGDQOVAFEQLHYRNNHTITIRPR
RIGGLWRAFAVSLDRALVAVGTVEGQAFRTPGAARDAVDAARSFIDARLKACRP
A"

397..555

/gene="RSp0644"

/note="synonym: RS05578"

397..555

/gene="RSp0644"

/function="miscellaneous; unknown"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Framed

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"

/protein_id="CAD17795.1"

/db_xref="GI:17431115"

/db_xref="GOA:Q8XS37"

/db_xref="TrEMBL:Q8XS37"

/translation="MERQLFAIVIAVAASAYSLARDLNDDEDFPPGQALVHGH
ATASAGR"

complement (575..1135)

/gene="RSp0645"

/note="synonym: RS05577"

complement (575..1135)

/gene="RSp0645"

/function="miscellaneous; not classified regulator"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Homology

predicted by Framed

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"

/protein_id="CAD17796.1"

/db_xref="GI:17431116"

/db_xref="GOA:Q8XS36"

/db_xref="TrEMBL:Q8XS36"

/translation="MKRISTFPDFVDMKKVAIKERSPAAAPQGAALSALDRTDRAIL
KWLQRDASINVALADKVSUSPPACLEKRRVERLKGCIIRGVVVALDEPKALDGLTLMVI
GVILDRSTPESFADFERAKVQKVPCCLECHVVVTGEFDFMLVTRTRDNDSYNRLHAEQLL
YLPGVGRIGRTMTLKLQVLTSTQLPIA"

1244..2260

/gene="RSp0646"

/note="synonym: RS05576"

1244..2260

/gene="RSp0646"

/EC number="4.1.99.4"

/function="miscellaneous; hypothetical/global homology"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Homology

predicted by Framed

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE
DEAMINASE (ACC DEAMINASE) PROTEIN"

/protein_id="CAD17797.1"

/db_xref="GI:17431117"

/db_xref="GOA:Q8XS35"

/db_xref="TrEMBL:Q8XS35"

/translation="MNLNKHPRHPLTTFGPTFIQPKRLSAHLGGKVELYAKREDCNSS
LAFGNKTRKLEYLVPEVLVAGGYDTLVSIGIQSNQTRQVAVAARAHGLKCVLQENW
VNYADAVYDRVGNIELSLRDLDAAGFDIGIRPSWEQAMEDVRAGGKPPFPIA
GCSBHVGLGLGVFAEVRQOEAEEFGFRFDYIVVCSVTGTOAGMVVGFADGRADR
VIGDASAKPQTRQILRIARDTAKLVELGRDITEDDVLDTYGGSEYGLPNSTGL
EAILRCARQEGMLTDPVTEGKSMGMDIRVRGGEFFPEGSRVLYAHLGGVPALNAYSFL
FRNG"

complement (2336..3508)

/gene="RSp0647"

/note="synonym: RS05575"

complement (2336..3508)

/gene="RSp0647"

/EC number="4.2.1.17"

/function="miscellaneous; hypothetical/global homology"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon usage

predicted by Homology

predicted by Framed

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"

/protein_id="CAD17798.1"

/db_xref="GI:17431118"

/db_xref="GOA:Q8XS34"

/db_xref="TrEMBL:Q8XS34"

/translation="WSALPKPAEPAAAPAEVLFHVNRVAITLNRPSALNALSHGM
VRQSALLAQCDDEPRIAAVVLRGNGEKGFCAGGVRALYRLARQGAHQPDGWLQFF
VDEYRLDYALHRFPKPVVALLDGTAMGGMLGQGRARLVVTERSRAMPETRIGFLP
DVGATHFLGAMPPEMALYIGLTGTLGADALHCLGADVCPAEWLASFEDLLRLSH
AGDPGLMQALRVFPEPCNIVPAAIARTTPWIVRYDFRSTIDRIVATLESRLERD
PAREARWLQATLDAMATHSPTMLHVTREALLRGQLTAEFCRMELGIVARAIIEGD
FCEGVRAHLVDKDRPGWAPATLAGVRPERVRHFLSSPWGAAHFLAHLGAQPQAA"

complement (3505..4305)

/gene="RSp0648"

/note="synonym: RS05574"

complement (3505..4305)

/gene="RSp0648"

/EC number="4.2.1.17"

/function="small molecule metabolism; degradation; carbon
compounds"

/note="Product confidence : probable

Gene name confidence : hypothetical

predicted by Codon usage

gene

CDS

gene

CDS

gene

CDS

predicted by Homology predicted by Frameb" /codon start=1 /evidence=not experimental /transl_table=11 /product="PROBABLE ENOYL-COA HYDRATASE PROTEIN" /protein_id="CAD17799.1" /db_xref="GI:17431119" /db_xref="GOA:Q8XS33" /db_xref="TrEMBL:Q8XS33" /translation="MTALDHLDTGTVALLTLRPPNATPEGLRLOLATVETLDADP RVALVITGGPKFSDADLNTFASGDVAREAAAREAGAFALONARPVIAIN GYAMGGLSCALACDIRIAERHAQLAVPETAVALCCGGTQTPLMVEGWAKRMIL TGERIDAQALRIGLVEEVDTGAAREALAMARRVTTLSPRAVTFSKALIHQARHG PTAALAVRERFVDFLGHDPDREGVNAFLERPPRWHTAARPEALP" complement (4302..5198) /genes="mmsB" /note="synonyms: RS05573, RSp0649" complement (4302..5198) /genes="mmsB" /BC_number="1.1.1.31" /function="small molecule metabolism; amino acid biosynthesis; isoleucine/valine" /note="Product confidence : probable Gene name confidence : putative Predicted by Codon usage predicted by Homology		Query Match Best Local Similarity 7.3%; Score 88.8; DB 1; Length 190050; Matches 291; Conservative 0; Mismatches 337; Indels 0; Gaps 0;	
QY	376 GATCTCGTCAATGATTGGAAGACGTTTCCGGTGTGCGTACGATGACATGCTCGGG 435		
Db	77778 GATCTCGTGGCGAATTCGCTTCCCGTTCGCGATCATCTGCCAGATGATGGAC 77837		
QY	436 CTGGATAAAGACACCATGAAATAATCTCTAGTGGCACAGCGGATGTCGCGATTTATC 495		
Db	77838 CTCGATATCGACGATGCCATGCGCTGGCGTGGGCGTGGAGCAAGCTCGCAAGGTGTT 77897		
QY	496 ACGAGTATCTCAATCTCTGAAGCGCGGCACATTCGTTATGTTGTCGACGACAGCTT 555		
Db	77898 GATCCGGCGCGATGAGCGCGATGCGTGTGGAGCAGACGCCGCCCTACGAGAACTC 77957		
QY	556 TCCCAATCTGATGCGCGTCAATTAAGAACGTCGCGTCAATCCGGGATCAGATTTAAT 615		
Db	77958 GCGCATATCTCAGAAGGTGATCGAGGCCCGCGCGCAGCCCGCACCGACTGAT 78017		
QY	616 TCGATCTATGATCTTCTGAATATGAGGATGCGCTGTGCGACAAAGATATCTCGCA 675		
Db	78018 TCGATCTGATGCGCGCGAGGAAAACGCGAGACGCTGACGACCATGAAATCGTGTCC 78077		
QY	676 CTGATCTTAATGTGCTGTAGCCGCAACGGACCGCTGATAGACGCTGCGACTGATG 735		
Db	78078 AACGTGATCTGTGTTCTATCGCGCGCCACGAGACCATCTCCACATGATCGGCAATGCG 78137		
QY	736 ATCTACATTTGCTCAACATCTCGAGCAGATGAATGATGTTTGTGCTGACCGTTCGTTA 795		
Db	78138 CTGATCGCTGTGATCGCATCGCAGACGCTGCACTGCTCAAGCGCGCGCTCGCG 78197		
QY	796 GTTCCGAGAGCCATTCGCGAGACATTCGCTTATAAACCGCGGTTTCACTGATTCGCGG 855		
Db	78198 ATGCCGAACGCGTGTGTAATTCCTGCGCTACGACGGGTGCGTGCAGGTACGATCCGC 78257		
QY	856 CAGCTGTCCNAGATACAGTGTGCGGGGTATGGAATCAAAAAGATACGATGTTTTT 915		
Db	78258 GCGGCGCTGGAGACATCGAAGTCGAAGGCGAGGTGCTGCCGCGCGCACGCGGTTC 78317		
QY	916 TGTATGATCGTGGCTGAACCGGACCTCGAGCATTTGAACAGCGCTGACGCTGTTTAT 975		
Db	78318 CTGATGCTGGCGCGCCATCGCATTCGGCCAGTTCACCGATCCGACCACTCGAT 78377		
QY	976 ATTATCGGGAAGATCTTGGTATCAAGA 1003		
RESULT 13 AE012559/c LOCUS DEFINITION Xylella fastidiosa Temeculal, section 7 of 9 of the complete genome. ACCESSION AE012559.1 GI:28057550 VERSION AE012559.1 KEYWORDS Xylella fastidiosa Temeculal SOURCE Xylella fastidiosa Temeculal ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xylella. REFERENCE 1 (bases 1 to 300885) AUTHORS Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorry, H., Tsai, S.M., Carrer, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Branco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Fenille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., Oliveira, A.R., Rosa Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P. TITLE Comparative Analyses of the Complete Genome Sequences of Pierce's Disease and Citrus Variegated Chlorosis Strains of Xylella fastidiosa JOURNAL J. Bacteriol. 185 (3), 1018-1026 (2003) PUBMED 12533478 REFERENCE 2 (bases 1 to 300885) AUTHORS Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, C.Y., Furlan, L.R., Camargo, L.E.A., da Silva, A.C.R., Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T., da Silva, F.R., Goldman, M.H.S., Goldman, G.H., Lemos, M.V.F., El-Dorry, H., Tsai, S.M., Carrer, H., Carraro, D.M., de Oliveira, R.C., Nunes, L.R., Siqueira, W.J., Coutinho, L.L., Kimura, E.T., Ferro, E.S., Harakava, R., Kuramae, E.E., Marino, C.L., Gigliotti, E., Abreu, I.L., Alves, L.M.C., do Amaral, A.M., Baia, G.S., Branco, S.R., Brito, M.S., Cannavan, F.S., Celestino, A.V., da Cunha, A.F., Fenille, R.C., Ferro, J.A., Formighieri, E.F., Kishi, L.T., Leoni, S.G., Oliveira, A.R., Rosa Jr., V.E., Sasaki, F.T., Sena, J.A.D., de Souza, A.A., Truffi, D., Tsukumo, F., Yanai, G.M., Zaros, L.G., Civerolo, E.L., Simpson, A.J.G., Almeida Jr., N.F., Setubal, J.C. and Kitajima, J.P. TITLE Direct Submission JOURNAL Submitted (17-DEC-2001) Instituto de Biociencias, Universidade de Sao Paulo, Rua do Matao, 277, Sao Paulo, SP 05508-900, Brazil FEATURES Location/Qualifiers 1..300885 /organism="Xylella fastidiosa Temeculal" /mol_type="genomic DNA" /strain="Temeculal" /db_xref="taxon:183190" /note="Pierce's disease strain" 87..1022 /gene="apbE" /locus_tag="PD1557" 87..1022 /gene="apbE" /locus_tag="PD1557" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="thiamine biosynthesis lipoprotein apbE precursor"			


```
OY 496 ACAGATATCTCTCAATCTCTGAAGCGCGGGGCACTTCTGTTATGGTGCAGCGAACAGCTT 555
Db 180936 GACCCCTGCGCCGATGTGCGCCGATGAATGGTGCAGCGAAGTCTGCTACGAGAACTA 180877
OY 556 TCCCAATACTTGTATGCGCGTCAATTAAGAACGTCGCGTCAATCCGGGATCAGATTTAAAT 615
Db 180876 GCGCATTAATTTCAAGGTTGATGAGTCGCGCGCGCGGAGCATGGAATGACTTGAAT 180817
OY 616 TCGATCCTATGTAATCTCAATATGAAGCATGCGCTGTCGACAAAGGATATACTCGCA 675
Db 180816 TCCATGTTGCTGCTGCTAGGAGGAGGCGCAGAACTGACGCATGATGAATCGTGTC 180757
OY 676 CTGATTTCTTAATGCTGTTAGCCGCAACGGAACCGGCTGATAGAGCTGGCACTGATG 735
Db 180756 AATGTAATTTGTTGTTATGTTGCTGCTAGGAGCACCACATCCAACTGATGGCAATGCA 180697
OY 736 ATCTACCATTTGCTCAACATCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTA 795
Db 180696 TTGATTCGGTTGATCGTCATCCGAGCAGCTTGCAAGCTCAAGAGTGAATTTGCTGCTG 180637
OY 796 GTTCCGAGAGCCATTCGGAGACATTCGCTTATAAACCCGCGGTTTCAGCTGATTCGCGG 855
Db 180636 ATGCCCAAGCGGATTTGGAATGTTTACGTTACGACGGATCAGTACAGTTTCAGATACGC 180577
OY 856 CAGCTGTCCCAAGATACAGTGTGCGGGGTATGGAATCAAAAAGATAGGATGTTTTT 915
Db 180576 GCGGCGATGGATGATGAGTATTAAGGTGATGTAGTCTGCTGCGCACGATTTGTTT 180517
OY 916 TGTATCATCGGTGCGCTACCGGGACCTGAAGCACTTTGAACAGCTTCAGCGTGTATAT 975
Db 180516 TTGATCTTGGTCTGCTACCGGTGATTCGCGCAATTCAGGATCCGACCAATTTGGAA 180457
OY 976 AT 977
Db 180456 AT 180455

RESULT 14
LOCUS AE017236/c 302325 bp DNA linear BCT 02-FEB-2004
DEFINITION Mycobacterium avium subsp. paratuberculosis str. k10, section 10 of
16 of the complete genome.
ACCESSION AE017236 AE016958
VERSION AE017236.1 GI:41396889
KEYWORDS
SOURCE
ORGANISM
Mycobacterium avium subsp. paratuberculosis str. k10
Mycobacterium avium subsp. paratuberculosis str. k10
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
avium complex (MAC)
REFERENCE 1 (bases 1 to 302325)
Li, L., Bannantine, J., Zhang, Q., Amons, A., Alt, D. and Kapur, V.
Direct Submission
Submitted (05-SEP-2003) Biomedical Genomics Center, University of
Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
LOCATION/Qualifiers
1. 302325
/organism="Mycobacterium avium subsp. paratuberculosis
str. k10"
/mol_type="Genomic DNA"
/strain="k10"
/sub_species="paratuberculosis"
/db_xref="taxon:262316"
156..2294
/locus_tag="MAP2433"
156..2294
/locus_tag="MAP2433"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA04750.1"
/db_xref="GI:41396889"
/translation="MRTARRRRTRADRNVLPGRVEIDNVQPVVSCGTYPAXAVVG
```

```
EVVPSASVWRGHEAATAATLVRYLGPAYPPVTETRRVKAQVAPVADREGVDQORI
KPLSLPMTGPEPYFHGQTFDQVGLMTFRDGLGWDGDIHSHWRHGLVAKLDAGQGETE
LNDLLVAGLEFERRATGVPRAARPEFLAAALADRTAGDPTVTRTALALAPEIEELAE
YPLRLLTRGEOYGVMDRPLARFGSWEMFPRSTCGDDGNGVHGTFTATAAALPR
IAAMPFDVYLPIHP IGVHRRGNNS TAGPTDVGS PWA IGSDEGGHDVHPDGLGT
IEDFAVARABELGMEVALDLALOCADPHPARHRWFTLPGDTIAYENPPPKY
QDIYELFNENDPAGLYDEVLVRVVRWHDGVKFFRVVDNHTFPDFFWALIAVKGID
PVLVLSAFTPVFRQNGTLKLGTSYTYFTWRTAKWELTFEFGDIALALPFRNNL
FYNTDIIHAILOHNGPGMFAIRAVLAATMGPAWGYSGYELFHRHVRGEESEYUNS
EYKELRPDRFAGALAEGLSEPTITQANTIRLHPALQQLRTIHFGHVDNDALLAYSK
FPDPTQDCLVVLVNTLNAFGPERATLFLDMAALGMEPYERFWVRDEITGQEFQWQANY
VRIDPAQAVAHVINMLPDEARMTLLRRR"
2329..4524
/gene="glgB"
/locus_tag="MAP2434"
2329..4524
/gene="glgB"
/locus_tag="MAP2434"
/codon_start=1
/transl_table=11
/product="GlgB"
/protein_id="AA04751.1"
/db_xref="GI:41396890"
/translation="MSHADQARTHLADPADLSRLVAGTHDHPGILGAHEVGDHTV
IKRAPHATEVAALVGGDRFAMQHIIESGLFAVLEFTNLIDIRLQITYPDSEPYVIVAD
AYRFLPTGEVDLHLFGGRHERLWEVLGAHPRSTADGVVTGVSFAVWNAKAGIS
LIGFNGTGTGTEAPMRVLGSSGWELFWDFPIGGLYKPKVHGADGVVTERADPMFA
TEVPHTASRVTLSSYTVAGDWMQRAARNPVFPMSTYEVHLGSWRPGLSVROLAR
ELTDVVEHGTHTVELLPVABHPFAGSGYQVTSYVAPTSRFGTDEPERALVDALHQA
GICVLDVWPAHFPKDAHALGRFDGTPLYEHSDDPKRGEQLDWGTYYVDFGRREVNF
NVALYMQEHPHIDGLRVDAVASMLYLDYSRPEGGTPTNLYGRENLEAVQLQENNA
TAKRSAPGITYAEESTSWPVTRPTNLGGLGFSMKNMGNMWHDTLDYISRDPYRSY
HHHEMTFSLYAFSENVLPLSHDEVHKGTLWGRMPGNHVKAGLSLLAYOWAH
PGKOLLFMQGEFQRAEWSERGLDWQDEQFSNGILRLVRDINDIYRSHPALWSQ
DTPDGYSWIDANDSANNVLSFLRYCKGDSVACVFNPAEHSYGLGLPSAGRWE
VLNTDATYNGSGIGNMGVDATEDPWHGRPASAVLVLPTLSALMLEPK"
complement(4550..5461)
/locus_tag="MAP2435c"
complement(4550..5461)
/locus_tag="MAP2435c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA04752.1"
/db_xref="GI:41396891"
/translation="NTRPRPSIGPALAGAVDLSGLKQRAAASGPAGGPAPADGGPGI
SAVTANFEAEVLIRSEEVVTVLLWSPEDACVOLLDLTSLAQQDNCKWASLATVNV
DVAPVQIFGVDAVPTVVALAGPISFQGNQADQLRGWLDQLLSATAGLKGAT
GSAESEAVDPELAAQQLEAGDFEAARTSYQAILDANPASTEAKGAIRKQIDFLTRAT
AQRPDVAVADAPGDIDAAFAAADVQILNQDVGAADFRLIALVRNTSGDRAVRTR
LVELDLPDPADEPVVAGRRNLNALLY"
complement(5652..6821)
/gene="fadA4"
/locus_tag="MAP2436c"
complement(5652..6821)
/gene="fadA4"
/locus_tag="MAP2436c"
/codon_start=1
/transl_table=11
/product="FadA4"
/protein_id="AA04753.1"
/db_xref="GI:41396892"
/translation="MIVAGARTPIGKLMGSLKDFASDLGAIAIAGALEKADVPADLV
EYVINGQVLTAGAGOMPARQAAGVIGDVPALSNKMLCSGLSDIALADOLIRAGE
FDVVVAGQESQWTKAPHLMLDSRAGYKGVGDVTLDMHAYDGLHDVFTDPMGALTEQR
NDVKFTREEDQYAAARSHQAAAWKQGVTFDEVVPVNIPOKRGDPLQFTDEGIRI
NTAESLAGLKPFRKDGTTAGSASQISDGAADVVMKAKAQQLGLSWLABIGHG
VVAGPDSTLQSQPANAIIKALREGISVDQLDQVINEAPAAVALASTRELGVNPDIV
NVNGGAIAGVHPIGMSGARIITLHAALSLRRSGVAVAAALCGAGCGQDALILRAG"
6915..7379
/locus_tag="MAP2437"
6915..7379
/locus_tag="MAP2437"
```



```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04754.1"
/db_xref="GI:41396893"
/translation="MTTDOVDARQVLAATAVTAIDHVGIAVADLDAALAWYHDHIGMI
LVHEVNEQGIQREAMLAVRGAPAGTAQQLMAPIDDSSTIAKFLDKRGPGIQQWAMR
LSDTLTLERLEQGVRLYDAPRRGTANSRINFHPKADAGGVLIELVEPA"
complement(7390..7686)
/locus_tag="MAP2438c"
complement(7390..7686)
/locus_tag="MAP2438c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04755.1"
/db_xref="GI:41396894"
/translation="WARRTPRRQRRLGPPPALRRVEVGPDPGHYEVKRPITAAARAVK
TVRCPCDHEIRSGAAHLVWVPEPTLGGAEARRHWHHTCGWAHRATRGPTRRWS"
complement(7710..8390)
/locus_tag="MAP2439c"
complement(7710..8390)
/locus_tag="MAP2439c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04756.1"
/db_xref="GI:41396895"
/translation="MSRVLRLVIAQCTVDYVGRLLTAHLPSARRLLLFKADGSVSVHADDD
RAYKLMWSPPCWPCREAGDAAPVWVENKAGEOLRITVEDIEHDSHDLGVDPLV
KDGVEAHLQALLAEHVQLLAGEVTLVRREYMTAIGPVDLLCRDRGGSVAVIKRGE
IDGVEQLTRYLELLNRDSVLAPVRGVFAAQIIPKQARTLATDRGLRCVTLDDYKVRGM
DSEYKRLF"
8427..10061
/locus_tag="MAP2440"
8427..10061
/locus_tag="MAP2440"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04757.1"
/db_xref="GI:41396896"
/translation="MSATQSTAQIRIGRVLEKITROSGRUPETPAYGSLLLGRVTSQH
RRRIQIIMTVWLGANLIGIGALILVIVAIAPSVFENDAPANITFVGPAVYVA
LAVGTWITRTVLAURWAIEBKPTPDERTNFIAPMRALVDVLVWAGFLLLTVL
RGLVDTMTFIRFLFGVFCGVLTACYLALFAIRPVAQALEARGPPPRRLTAGIMG
RTWYVFLSGVFPVIGIALALFEIWLRLTETEFAGVLIIVSTAALIIFGCLLMILA
WLTATPVRVRAALKRVERGDLRVFDGTGELGELQRFNAMDGLRERVRDLF
GRHVGEVLAALAEEREPKLGSRERHVAVFIDVGTOLATVRRPPAEVVSVLNPFGI
VVEVDHRCGLNKEGDATLAI FGNPHLDCPEDAALTAARAIADRLANENPQCEAG
IGVAGQVAGNVAGAREFEYTVIGPVNEAARLCELAKSHSRLLATGDATGEGASEK
ERARSLGDTVTLRHERPTRLASPAAGADRPAQTD"
complement(10525..11787)
/locus_tag="MAP2441c"
complement(10525..11787)
/locus_tag="MAP2441c"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04758.1"
/db_xref="GI:41396897"
/translation="MPARSMVWPHGPVHAQAGTRCYRMTAALCGAGLASFAAMYCS
QALLPALSHYRIGPATAALVSLTIGALALSIIPASVLSERYGRIRVMLSIGVASSV
IGLLFPFSLGVLLFGRAAQGVAGALGIPAVAMALLAEVEDASSLGSAMGRYIAGTTI
GGLAGRIVPSVQVQTVVALLACSLITLAGTAVFAVLVPRSRFTPKPASVRAALR
NLAGLRNPVLAKLFAVGFLVWGGFVTVVNYLRAARPFGLASVVGLLFLLVLG
TGTSVVAGLRADRRGRPLVIGALPIAVAGLLITVPATIAAIVAGVFTGGFFAHT
VASGWGAVAQDRDRAEASALYLSYVLGGSVAGAGGVLYVGGSATVCFVVVLLMA
GAALVALLVRDNGFRIGRRVTVSVASK"
11854..12738
/locus_tag="MAP2442"
11854..12738
/locus_tag="MAP2442"
```

```
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAS04759.1"
/db_xref="GI:41396898"
/translation="MAVDGYEMFTLAEQLHVTAABQLHVAQPTLTRMLARLEORLQ
VALPDRHRRLSLNTYGRIFYEHARRAQLEDSARRAIALADLAPVAGVIRLAYLSFG
STVPLRLTARFKESSPRVTFTLEKGAESIADRVRSGGVDIGVSGPRVPVTLAWRSL
FRQLGVAVPDGHRFARGAANVMDLADPEFTWHPHFGMRLELDELCAAAQFRPRVA
LQENLTAAASLVAAGLGISLVPIDGSSYFSGSVSKPLADADAYRDVGNWDSGRPLP
LSARDFIAAASAVRGGW"
12825..14348
/locus_tag="MAP2443"
12825..14348
/locus_tag="MAP2443"
/codon_start=1
/transl_table=11
/product="Alka"
Query Match 7.0%; Score 85.2; DB 1; Length 302325;
Best Local Similarity 49.8%; Pred. No. 3.7e-14;
Matches 216; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
QY 552 GCTTTCCCAATACTGATGCGGTCAATTAAAGAACGTCGCGTCAATCCGGGATCAGATTT 611
Db 252449 GCTGTGCGACTACTTCGGCCGATCCTGGCGCGCGCGGAGAGCCCAAGGACGATCT 252390
QY 612 AATTTTCGATCTTATGTAATCTTCTGAATATGAAGGCATGGCGTGTTCGGAACAAGGATATCT 671
Db 252389 GATCAGCGGCTGGCCAGCGGAGATCGACGCGGAGAACTGGAAGACGAGGAGATCTA 252330
QY 672 CGCATCGATTTTAAATGCTGTGTAGCCGCAACGAAACCGGCTGATGAACGCTGGCACT 731
Db 252329 CTCCTTCTCGCGCTGCTGCTCGCGCGCGGGTGGAGACGACTACCGCGCGCTGGGCGAG 252270
QY 732 GATGATCTACCATTTGCTCAACAATCCTGAGCAGATGAATGATGTTTTCGCTGACCGTTC 791
Db 252269 CTTGCTGCTTGGCTGCTGCTCGGATCCGAGCAGTTGATGTCATCCGCGGGAGCCGCTC 252210
QY 792 GTTAGTTCGAGAGCATTGCGGAGACATTTGCGTTTATAAACCGCGCGTTTCACTGATTC 851
Db 252209 GCTGCTGCGCAGGCCATCGAGAGGGCGTGCCTGGGAGCCCGCTCTGACCATCAC 252150
QY 852 GCGGAGCTGTCCCAAGATACAGTGGTGGCGGTTATGGAATCAAAAAGATACGATTTGT 911
Db 252149 CCGGCTGGCCACCCCGCAGACCGAACTCGCGGGGTGCCGATCCCGCGCGGTCCACG 252090
QY 912 TTTTGTATGATCGGTGGCTAACCGGACCCCTGAAGCATTTGAACAGCCTGACGCTGT 971
Db 252089 GATGCGGATGCTGGGCGCGCCCAACCGGAGAGGACCGCTACCCCGACCCGACAGGTT 252030
QY 972 TAATATTCATCGG 985
Db 252029 CGACATCTTCGCG 252016
```

RESULT 15

```
AF015825
LOCUS AF015825 35739 bp DNA linear BCT 07-MAY-1998
DEFINITION Bacillus subtilis 168 cot-rapA region sequence.
ACCESSION AF015825
VERSION AF015825.1 GI:2612880
KEYWORDS
SOURCE
ORGANISM Bacillus subtilis
Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 35739)
AUTHORS Rivolta,C., Soldo,B., Lazarevic,V., Joris,B., MaueI,C. and
Karamata,D.
TITLE A 35.7 kb DNA fragment from the Bacillus subtilis chromosome
containing a putative 12.3 kb operon involved in hexuronate
```



```
/translation="MAGKKTITINGVEMAESEBQTVLQLNNSIEVQVCYHPSLGP
1ETDCTIVSINGELKRSABLDGDVIDITLSPDVKKQVIMGDKILYNHLEYICTVC
DYNNGCEIHTNVEKMKINHOSIPFDHKPYHKDESHPFYRVDPDQICILGRCVEACQD
VOVTEFLIDWERKEPRVITWNDVPINESSCVSGHCSTVCEPCNMEKMGEGAGYL
TOINNETIAPMEITKGVTGVSILAIASDMESAMRDERIKKTKVTCTYCGVGSFDV
WTKGHDILUKVEPQEEBANGISCTVKFGFMDVFNSEBELTKPLIREGDHFRFAEWEE
ALLIASKFTELKEAFGPDLSLAFTTSSKCTNEESYLMQKARGVIGTNNVDCSRYCQ
SPATAGLFRITGVGSDSGSITDIAQADLVLIIGSNTSESHPVLSRIKRAHKLRQKV
IVADIRKHEMAERSLFPQPRAGSDIVMLNAIAKVLIEGKADERFLRERVNGRDEYV
KSLAPYTYEABEKGIDQETLIQMAEMIQDASVCALWAMGVTOHIGGSDTSTAISN
LLLVNTGNYKPGAGSPLRGNHNVGASDFGSDPRLPGYEVKTDEQVRQKTERVMGV
PLPKPFGMTHNEMIKIHSQOLKAMYVRKEEMGLVDSNINHVAAYEKLDFFVODIF
LSRTAEFADVPLPASPSLEKEGTFTNTERRIQRLYQVFEPLGESKPDQIIMEVANKL
GAGWLEHPADIMEEAAKLPIYAGVTYERLEHFNGLVKNVADCKDPLLFTFERPF
PDGKAILYPOVOTPEKPERCEEVDIHVNGRLLLEHFEHGNLTVKSGISEKTPVEFLEI
SPELAEARGIQDGLTVRLTSPFGNVKVKCLITDRVKGEVILPMDNSEAAINLLTGS
HADKDTDTPAYKETSAKMEILKHDGIDSLPKINHRNGNPQPOIGVQVHKWKARDYIF
PGDAVKRGMGNHG"
7762..8373
/gene="y1gd"
RBS 7762..7769
/gene="y1gd"
CDS 7780..8340
/gene="y1gd"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC46313.1"
/db_xref="GI:2612889"
/translation="MAKAIKRIQKIEVTEEDQQRKDLREIEDALIDHKEAILETLHML
GHMNERGLPLLRGLFGQDKVLDILVKKADTEANTLKNLLILFLGLMDLVQKLE
PLILKNAGVASAVEQKSEKTYGYFDIIRSLKDPPEINKSITLLFSFLKGMGQDTKEL
ERTQPPHQKHQEPREKRNKRD"
8347..8373
/gene="y1gd"
RBS 8525..9223
/gene="y1ha"
RBS 8525..8530
/gene="y1ha"
CDS 8537..9178
/gene="y1ha"
```

stem_loop

```
Query Match 6.5%; Score 78.8; DB 1; Length 35739;
Best Local Similarity 45.4%; Pred. No. 2.3e-12; Indels 6; Gaps 1;
Matches 327; Conservative 0; Mismatches 387;

QY 270 AATTGTAGTGAAGCTTTATCGGTGACGCACTGGATCATCTGTCTCCATTGATTAACA 329
Db 10864 AGTCGTGAACAAGCCCTTTACTCCGCGCGTGATGAAGCAATGGAAACCGAGAATTCAGA 10923

QY 330 AAATGCGAGAAAACCTGTTAGCGCCTTATCTTGAAGAGGGAAAAGTGATCTCGTCAATGA 389
Db 10924 AATCACAGATGAACCTGATCAAAAAATTCAGGGCGCGAGTGAGTTTGACCTTTGTTCACGA 10983

QY 390 TTTTGGAAAAGAGCTTTGCGGTGTGCGTCACGATGGACATGCTCGGCTCGATAAAAAGAGA 449
Db 10984 TTTTTCATACCGCTTCCCGGTTATGTGTATCTGAGCTGCTGGGAGTCCCTTCAGCGCA 11043

QY 450 CCATGAAAAAATCTCTGAGTGGCACAGCGGAGTTGCCGATTTTATCACGAGTATCTCTCA 509
Db 11044 TATGGAACAGTTTAAAGCATGGTCTGATCTTCTGGTCAGTACACCGAAGATAAAGTGA 11103

QY 510 ATCTCTGAA-----GCGGGGCAATTCGTTATGGTGACGGAACAGCTTTCCCAATA 563
Db 11104 AGAAGCTGAAAAAGCCCTTTTGGAGAACGAGATAAGTGTGAGGAAGAACTGGCCCGCTT 11163

QY 564 CTTGTATGCGGTCATTAAGAACGTCGCGTCATCCGGATCAGATTTAATTTGATCCT 623
Db 11164 TTTTGGCGGCATCATAGAAGAAAAGCGAAACAAACCGGAAACAGGATATTTTCTATTTT 11223

QY 624 ATGTACTTCTGAATATGAAGGCATGGCGGTGTGCGACAAGGATATATCTCGCACTGATCT 683
Db 11224 AGTGAAGCGGAAGAAACAGCGGAGAAGCTGTCCGGTGAAGAGCTGATTCGTTTTGCAC 11283
```

Search completed: December 10, 2004, 16:12:22
Job time : 5416.29 secs

```
QY 684 TAAATGTGCTGTTAGCCGCAACGGAACCGGCTGATAAGACGCTGGCACTGATGATCTACCA 743
Db 11284 GCTGCTGCTGCTGCGCGGAATGAACCACTACAAACCTGATTTCAAATGCGATGTACAG 11343

QY 744 TTTGCTCAACAATCCTGAGCAGATGAATGATGTTTGTGGCTGACCGTTGTTAGTTCCGAG 803
Db 11344 CATATTAGAAAACGCGCAGCGCTTTACGAGGAACTGCGCAGCCATCCTGAACGTGATGCCTCA 11403

QY 804 AGCCATTGCGGAGACATTTGCGTTATAAACCGCGGTTTCAGCTGATTCGCGGCGAGCTGTC 863
Db 11404 GGCAGTGAGGAAGCCTTTGCGTTTCAGAGCGCGCGCCCCGGTTTTGAGGCGCATTCGCCAA 11463

QY 864 CCAAGATACAGTGGTCCGCGGTATGGAATCAAAAAAGATACGATTTGTTTTTGTATGAT 923
Db 11464 GCGGATACGGAGATCGGGGGCCACCTGATTAAGAAAGTGATAGTGGTTTGGCGGTTTGT 11523

QY 924 CGGTGCGCTAACCGGGGACCTTGAAGCATTTTGAACAGCGCTGACGTTGTTTAATATTCATCG 983
Db 11524 GGCATCGGCAAAATCGTGATGAAGCAAAAGTTTGACAGACCGCACATGTTTGTATATCCGCG 11583
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 11:14:31 ; Search time 672.942 Seconds
(without alignments)
9477.854 Million cell updates/sec

Title: US-10-627-124-1
Perfect score: 1215
Sequence: 1 atagaccaatcgattaaatt.....tcgttcgcttgacggggca 1215

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1215	100.0	1215	12	ADJ62038
2	617.2	50.8	1208	6	ABK74890
3	247.4	20.4	1224	10	ACF70115
4	247.4	20.4	110000	10	ACF65385.2
5	247.4	20.4	110000	10	ACF67367_31
6	82	6.7	1204	3	AA95667
7	81.4	6.7	1188	9	ACF06114
8	78.8	6.5	1191	3	AA95666
9	78.8	6.5	1221	3	AA95668
10	77.2	6.4	1221	3	AA95669
11	72.4	6.0	966	6	ABK74899
12	67.8	5.6	110000	4	AAI99682_39
13	67.8	5.6	110000	4	AAI99683_39
14	64.8	5.3	1213	6	ABK74891
15	61.8	5.1	110000	4	AAI99682_08
16	61.8	5.1	110000	4	AAI99683_08
17	59.2	4.9	25681	10	ADJ53165
18	55.6	4.6	1200	8	ACA38018
19	55.6	4.6	84428	12	ADM45913
20	55.2	4.5	1233	10	AAD55813
21	55.2	4.5	60196	10	AAD55810

22	55	4.5	1155	6	ABK74894
23	54.2	4.5	36538	10	ABV75558
24	53.6	4.4	1215	9	ACF06133
25	53.6	4.4	110000	4	AAI99682_08
26	53.6	4.4	110000	4	AAI99683_08
27	53.4	4.4	1400	2	AAQ14548
28	52.8	4.3	5880	4	AAI68687
29	52.8	4.3	11355	4	AAI68692
30	51.6	4.2	2000	8	ADA71938
31	50	4.1	17596	12	ADI14148
32	50	4.1	47981	4	AAF30757
33	49.6	4.1	1191	9	ACF06115
34	49.6	4.1	6885	2	AAI70153
35	48.2	4.0	9024	12	ADI14147
36	47.6	3.9	1209	8	ABX56043
37	47	3.9	1230	10	ADC36274
38	47	3.9	1449	10	ADC36279
39	46.6	3.8	1290	5	AAH65582
40	46.6	3.8	1407	4	AAF71749
41	46.6	3.8	1414	10	ADD13338
42	46.6	3.8	2000	8	ADA71938
43	46.6	3.8	34980	5	AAH68525
44	46.2	3.8	1248	9	ACF06136
45	44.8	3.7	1227	10	ADC36139

ALIGNMENTS

RESULT 1
ADJ62038
ID ADJ62038 standard; DNA; 1215 BP.
XX
AC ADJ62038;
XX
DT 06-MAY-2004 (first entry)
XX
DE Bacillus subtilis cypX encoding DNA SEQ ID NO:1.
XX
KW heterologous biological substance; Bacillus; cypX; red pigment; gene; ds.
OS Bacillus subtilis.
XX
FH Key Location/Qualifiers
FT CDS.
FT 1..1215
FT /*tag= a
FT /product= "cypX"
XX
PN WO2004011609-A2.
XX
PD 05-FEB-2004.
XX
PF 25-JUL-2003; 2003WO-US023398.
XX
PR 26-JUL-2002; 2002US-0398853P.
XX
PA (NOVO) NOVOZYMES BIOTECH INC.
XX
PI Tang M, Sloma A, Sternberg D, Behr R;
XX
DR WPI; 2004-143839/14.
DR P-PSDB; ADJ62039.
XX
PT Producing a heterologous biological substance comprises cultivating
PT pigment-deficient mutants of Bacillus cell that directs synthesis of the
PT heterologous biological substance and has a modification of the cypX and
XX ymc genes.
PS Example 1; SEQ ID NO 1; 62pp; English.
XX
CC The present invention describes a method for producing a heterologous
CC biological substance comprising cultivating a mutant of a parent Bacillus
CC cell in a medium suitable for the production of a heterologous biological

Handwritten signature

(NOVO) NOVOZYMES AS.

Berka R, Clausen IG;

WPI; 2002-416684/44.

Monitoring differential expression of several genes in first Bacillus cell relative to expression of same genes in one or more second Bacillus cells, by using substrate containing Bacillus genomic sequenced tag array.

Claim 4; SEQ ID NO 2181; 200pp; English.

The invention describes a method of monitoring differential expression of genes in a first Bacillus cell relative to expression of the genes in other Bacillus cells, comprising hybridising labelled nucleic acid probes isolated from Bacillus cells to a substrate containing array of Bacillus genomic sequenced tags (GST), examining the array, and determining relative gene expression by an observed hybridisation reporter signal of a spot in the array. The method is useful for measuring the expression of genes in a first Bacillus cell relative to expression of the same genes in one or more second Bacillus cells. The method is useful for monitoring global expression of several genes from a Bacillus cell, discovering new genes, identifying possible functions of unknown open reading frames and monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which Bacillus cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

Sequence 1208 BP; 318 A; 321 C; 307 G; 262 T; 0 U; 0 Other;

Query Match 50.8%; Score 617.2; DB 6; Length 1208;
Best Local Similarity 70.5%; Pred. No. 1.2e-186;
Matches 853; Conservative 0; Mismatches 353; Indels 4; Gaps 2;

PA	1	ATGAGCCAAATCGAATTAATTTGTTAGTGTGCTTTTGTGATCAATTTCAAAACAAATCCATAT	60
XX			
PI	1	ATGATCAATCGCTAAACAAATTTACGCGTGTGTGAGAAATATCAAGAAATCCGTAT	60
XX			
DR	61	GCTTATTTTCAACTCGCGGAGGAGATCGGTTTCATATGAAAGATCGATAGACAGT	120
XX			
PT	61	CAGTATTTCTCATACCTTCGGGAATCTGATCCCGTCCATTTATGAAGAAATCGCTGGACAGC	120
XX			
PT	121	TATTTTATCAGCGCGTATCATGATGTCGGTATATCTTTCAGCATCCGGATATCTTCACG	180
XX			
PT	121	TATTTTATCAGCGCGTATCATGATGTCGGTATATCTTTCAGCATCCGGATATCTTCACG	180
XX			
PT	181	ACGAAAT--CACTTGTGTAGCGTCCGAAACAGTCAATGCGAGGCCCTGTGTCGCCCAAA	238
XX			
PT	181	ACGATTTGCTGTGCGCAACGCGCGAGCGGCTCATGCGCGGACCTGTGTCGCCCAAA	240
XX			
PT	239	TGATGGAAAGAAACACTCTGCGCAAGAAAGAAATTTAGTGAGAAAGCTTTATCGGTGACG	298
XX			
PT	241	TGAAGGCAAGAGACACAGCGCAAGAAAGAGAAATCGTTCTGCGCGCTTTATCGGGGAAT	300
XX			
PT	299	CAGTGCATCATCTGCTCCATTTATTAACAAATCGAAGAAATTTAGCGCCTTATC	358
XX			
PT	301	CCCTTGCATCACTTCACACCGCTCATCAAGAAATTTGCGGCTTTTGGCGCCGAC	358
XX			
PT	359	TTGAAAGAGGAAAGTGTCTCGTCAATGATTTTGGAAAGAGCTTTGCGGTGTGCGTCA	418
XX			
PT	359	GTGGAAAGGGCGGATCGACCTTGTCAATGATTTTGGCAAAACATTCGCGGTTTGGTGA	418
XX			
PT	419	CGATGACATGCTCGGGCTGGATAAAGAGACCATGAAATAATCTCTGATGTGCAACGCG	478
XX			
PT	419	CGATGACATCTTTAGGTTTGGACAGAGACGACCAACAAACGGGTGAGAAATCGGCACAGCG	478
XX			

QY	479	GAGTTGCCGATTTTATCAGAGTATCTCTCAATCTCTGAAGCGCGGCAATTCGTTAT	538
DB			
QY	479	GOGTCGCGATTTCAATACCAGTTTGAATCAGGCGCTGAGGACCGGGAGCATTCCTCA	538
DB			
QY	539	GGTGAGGAAACAGCTTTCCCAATATTGATCCCGTCAATTAAGAAACCTCGGTCAATC	598
DB			
QY	539	AATGAGTGAACAGCTCGCTGAGTATTTGAATCCGATATTCGAGGAAAGCGGCAAAATC	598
DB			
QY	599	CGCGATCAGATTAATTTTCGATCTATGATCTTCTGAATATGAAGGCATGCGGCTCGG	658
DB			
QY	599	CGGACATGATTTAATATCCATCTTTCGACATTCGGAATACGAGGAGTGGCGATCTG	658
DB			
QY	659	ACAAGATATATCGCACTGATTTCTTAATGTGCTGTAGCCGCAACCGGCTGATA	718
DB			
QY	659	ACCGGACATACCGCGCTGATTTCTCAATATATTTGTTCCGCGCACAGAACCGGCGACA	718
DB			
QY	719	AGACCTGGCACTGATGATCTACCAATTTGCTCAACAATCTCGACAGATGAATGATTT	778
DB			
QY	719	AGACGCTGCAATTAATGATTTATCATTTTCTTCATCATCCGATCAGATGAATGATGTC	778
DB			
QY	779	TGGCTGACCGTTCTGTTAGTTCCGAGAGCCATTTCCGAGACATTTGCGTTATAAACCCCGG	838
DB			
QY	779	TGGAAGACCGACGCTTCTCCCTCAGGCAATCGCAGAGAGCTGCGCTACAGCCCGCGG	838
DB			
QY	839	TTTACGCTGATTTCCGCGGACGCTGTCCAAAGATACAGTGTGCGCGGTATGGAAATCAAA	898
DB			
QY	839	TGCAGCTCATCCGCGCCAGCTTTTCAAGACCGCGAGATTGCGCGAGTCGAGCTAAAG	898
DB			
QY	899	AGATACGATTTGTTTGTATGATCGGTGCGGCTAACCGGACCTGAAGCATTTGAAC	958
DB			
QY	899	AAGGACGACTGTATTTTGCATGATAGCGCGGCAATCCGATCCTGAAGCGTTTCGAGG	958
DB			
QY	959	AGCCTGACGTGTTTAAATATTTCATCGGGAAGATCTTGGTATCAAGAGCGCTTTAGCGCGG	1018
DB			
QY	959	ATCCCGACAGTTCAACATTTACCGCAGACGCTGGAAGTCAAAAGCGCATTCAGCGGCG	1018
DB			
QY	1019	CGCCCGGCATCTCGCTTTCCGATCCGCAATTCATAACTGTGTAGAGCAGCTTTTGCCA	1078
DB			
QY	1019	CAGCCAGGCATCTCGATTTCCGTTTCCGCTTCCAGGCGTCCACAACTGTGTAGAGCGGGTTGCGA	1078
DB			
QY	1079	AAAAAGAAATCAAAATTTAGCTAATATTTGCTGATGAAGATCGGATATCAGATTAG	1138
DB			
QY	1079	AGACCGAAATCAGCTTTGCGCAATATCTGCTTGTATCAGCTGAAAAATATCCGGCTGG	1138
DB			
QY	1139	AGGAAGATTTTGTATGCTGAGTCCGCTCTGTATACAGCGGACCTGTTTCACTTCTCG	1198
DB			
QY	1139	AGGAAGATTTTGTATGCTGAGTCCGCTCTGTATACAGCGGACCTGTTTCACTTCTCG	1198
DB			
QY	1199	TTGCGGTTTGA	1208
DB			
QY	1199	TCCGGTTTGA	1208
DB			

RESULT 3

ACF70115

ID ACF70115 standard; DNA; 1224 BP.

XX ACF70115;

XX ACF70115;

DT 20-NOV-2003 (first entry)

XX Photorhabdus luminescens nucleotide sequence #8582.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

XX detection; food; gene expression; plant; animal; microorganism; toxin;

XX antibiotic; bioplastic; virulence factor; disease model; plague;

XX whooping cough; gene; ds.

XX Photorhabdus luminescens.

XX WO200294867-A2.

PN

XX

XX

PD 28-NOV-2002.

83246 AATTCTATAAAATCCTTACGATTTTTCAGACATTTTTCATATAAACAAGATTTAGTTTATT 83305
QY |||||
101 ATGAAGAGTCGATAGACAGTTATTTTATCAGCGCTATCATGATGTCGCCCTATATCTCTC 160
Db |||||
83306 TTGAACAATCCAAATAGCTATTTTATTTGGCAATATAGAGATGTTGAGCTATTTTAA 83365
QY |||||
161 AGCATCCGGATATCTTCACGAGAAATCACTTGTGTGAGCGTCCGAAACCAAGTCATCGGAG 220
Db |||||
83366 AATCTCTCA---ATTTTCAATATAAGCCATTAATGCACTTGCAGTCCGAAACGGTTATGGGG 83422
QY |||||
221 GCCCTGTGTCGCCCAATGATGTAAGAGACACTCTGCCAAAGAAAGAAATTTAGTCA 280
Db |||||
83423 ATCGCGCTTCTGCTCAATAGGAAGGGAAGAGATGCGGTAAACGAAATTCATAATGC 83482
QY |||||
281 GAAGCTTTATCGGTGACGCACTGGATC---ATCTGTCTCCATTTGATTAACAAAATGACG 337
Db |||||
83483 AAGGACTTTCAGAGATTTATTTTATCGCTACTACGAACCCATGATTCGAAGATTTACTG 83542
QY |||||
338 AAAAATGTTAGCGCTTATCTTGAAGAGGAAAGTGAATCTCGTCAATGATTTTGGAA 397
Db |||||
83543 AAGATCTCTTCAACCTTATATGAAAGAAAGAAATATAGATATTTGTAATGACTTTGGCC 83602
QY |||||
398 AGACGTTTGGGTGTCGTCACGATGGACATGCTCGGGCTGGATTAAGAGACCAATGAA 457
Db |||||
83603 GCGATTTATCCGCTATTAGTGACATTAAGATTTCTCGGACTGCGAAGTGAATACTATCGTG 83662
QY |||||
458 AAATCTCTGAGTGGCAGCGAGTTCGCGATTTTATACGAGTATCTCTCAATCTCTCTG 517
Db |||||
83663 ATATAGCTGATGCGACAGGGTATCGCCAGTTTATATACCGAGTTTGACCAACAGAAC 83722
QY |||||
518 AAGCGCGGCAATTCGTTTATGGTGAGCGAAACAGCTTTCCAAATACTTGAATGCGGGTCA 577
Db |||||
83723 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAATTAATTCGCTTACTTAAAGCCTATAA 83782
QY |||||
578 TTAAGAACGTCGCTCAATCCGGATCAGATTTAATTTTCGATCTCTGATCTCTGAT 637
Db |||||
83783 TAGATCAGAGCGGGTATATCCAGTAAAGCAATTTATCTATCTATCTG----- 83831
QY |||||
638 ATGAAGCATGCGCTGTCGGAAGGATATCTCGCACTGATCTTAAATGCTGTGTAG 697
Db |||||
83832 -TCAGATACGGATGTCATGAGTGAATTTACAGCATGTGTTTAAACATCTCTATTAG 83890
QY |||||
698 CGCAACCGAAGCGGTGATAGACGCTGGCTGATGATCTACCAATTTGCTCAACAATC 757
Db |||||
83891 CGGCTACAGAACTGCTGCAAGATATTAGCAATGATGCTTAACTAAATATCTAATC 83950
QY |||||
758 CTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGCGATTCGGAGA 817
Db |||||
83951 CGAGTATGCTTGATGTAGTTCTCAAGATCGCAGCTTAGTTCGAGATGCAATTTGAAGAA 84010
QY |||||
818 CATTCGGTTTAAACCGCGGTTTCAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGG 877
Db |||||
84011 CTTTACGTCGATCATCCCGTACAACTCATTTCAAAGAGGGAGTGAAGAGCTCACTA 84070
QY |||||
878 TCGGCGGTATGAAATCAAAAAGATACGATTTGTTTTGATGATCGGTGGGCTTAACC 937
Db |||||
84071 TTTCTGTTATCGATATACCTAAAGCGCTGTAGTATTTGTAATGATGGCGAGCTAATC 84130
QY |||||
938 GGGACCTCGAAGATTTGAACGCTGAGTGTGTTAATATTCATCGGGAAGATCTTGGTA 997
Db |||||
84131 GTGATCCATCGGTTTTTCAAAACCAATGAATTTGATCTATATCGAAGAAAAATACCA 84190
QY |||||
998 TCAAGAGCGCTTTAGCGCGCGCCGCTCTCGCTTCGATTCGATTCGCGATTCATACT 1057
Db |||||
84191 CTTTCTCCAGAAAGCAATTCGAAAGACATTTAGCTTTTGGTGAGGTACTCATGCCT 84250
QY |||||
1058 GTGTAGGAGCAGCTTTTGCMAAAACGAAATCGAAATTTAGTAAATTTTGTGCTGATA 1117
Db |||||
84251 GTGCTCGGCTGATTTTCTCAGTCAGTGGAGTTTCATCCATATATTTTGGATC 84310
QY |||||
1118 AGATCGGAATATCAGATTAGAGGAAGATTTTGTGTTATGCTAGTCCGCTGTGATACAC 1177
Db |||||
84311 TCTGCAATATTTACGTTTGTGTCGATCACTATCATTAATCAAGAAACAGGCGTGTATACAC 84370

QY 1178 GGGAGCCTGTTTCACTTCTCTGTTGCGTTGA 1208
Db 84371 GAGGACCTCTAAACTCCCTTTGAGTTTGA 84401

RESULT 5

ACF67367_31
Continuation (32 of 57) of ACF67367 from base 3100001 (Photorhabdus luminescens nucleotic
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	564894

Query Match 20.4%; Score 247.4; DB 10; Length 110000;
Best Local Similarity 52.7%; Pred. No. 5.7e-67;
Matches 617; Conservative 0; Mismatches 536; Indels 18; Gaps 3;

QY 41 AATTTCAAAACAATCCATATGCTTATTTTTCACAACTGGCGGAGGAAGATCCGGTTCAAT 100
D8 88380 AATTTCATAAAAATCCTTACGATTTTACGACATTTTGCATATAACAGATTTAGTTTAT 88439
QY 101 ATGAAGAGTCGATAGACAGTTATTTTATCAGCGCGTATCATGATGTCCTATATCCTTC 160
D8 88440 TTGAACAATCCCAAAATAGCTATTTTATTTGGCAAAATATGAGGATGTGACGCTATTTTAA 88499
QY 161 AGCATCCGATATCTTCACGACGAATCACTTGTGAGCGTCCGACCCAGTCATCGGAG 220
D8 88500 AATCTCTCA---ATTTTCAATACTAAGCCCAATTAACATGCACTTCCGACCCGGTTATGGGG 88556
QY 221 GGCCTGTGTCGCCCAATATGCAATGGAAGAAAGACACTCTGCCAAAGAAAGAAATTTAGTGA 280
D8 88557 ATCGCGTCTTCTCTAATGGAAGTGAAGAGGATGCGGTGTAACGAAATTCATATGC 88616
QY 281 GAAGCTTTATCCGTGACGCACTGGATC---ATCTGTCTCCATTTGATTAACAAAATGCAG 337
D8 88617 AAGGACTTTCAAGAGATTTATTTAATCGCTACTACGAACCCATGATTCGAAAGATTACTG 88676
QY 338 AAAACTTGTAGCGCTTATCTTGAAGAGGGAAGAGTATCTCGTCAATGATTTTGGNA 397
D8 88677 AAGATCTCTTACACCTTTATATGGAAGAAAGGAAATATAGATATTTGTAATGACTTTGGCC 88736
QY 398 AGACGTTTCGCGTGTGCGTCACGATGGACATGCTCGGGCTGGATAAAGAGACCATGAA 457
D8 88737 GCGATTATGCCGTATTAGTGACATTAAGATTCTCGGACTGCGAAGTGAATACTATCTG 88796
QY 458 AAATCTCTGAGTGGCAGCGGAGTTGCGGATTTTATCAGAGTATCTCTCAATCTCTG 517
D8 88797 ATATAGCTGAATGGCACAAGGATGATCGCCAGTTTATATCCAGTTTGACCAACAGAAC 88856
QY 518 AAGCGCGGCACATTCGTTATGCTGAGCAGCAACAGCTTTCCCAATACITGATCGCGTCA 577
D8 88857 TAGAAAAAATGCACAGTCTTGAATGTAGTCAGAAATTAATTCGCTTACTAAAGCCCTATA 88916
QY 578 TTAAGAGACGTCGCGTCAATCCGGATCAGATTTAATTTCCGATCTATGCTACTTCTGAAT 637
D8 88917 TAGATCAGAGACGGGTAAATCCAGTAAGGACATTAATATCTATTTCTG----- 88965
QY 638 ATGAAGGCATGGCGTGTGTCGCAAGGATATACTCGCACTGATTTCTTAATGTGCTGTAG 697
D8 88966 -TCAAGATACGCGATGTCCATGAGTGAATTTACAGCACTGTGTTTAAACATTTCTATTAG 89024
QY 698 CCGCAACGAAACGGCTGATAGACGCTGGCACTGATGATCTACCAATTTGCTCAACAATC 757
D8 89025 CGGCTACAGAACTGTGCAAGAGATATTAGCAATGATGCTTAATCACTTAATCTTAATC 89084
QY 758 CTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGCCATTGCGGAGA 817
D8 89085 CGAGTATGCTGATGTAGTTCTCAAGATCGCAGCTTAGTTCGAGATGATTTGAAGAA 89144
QY 818 CATTCGCTTATAAACCGCGGTTTCAGCTGATTTCCGCGGACGCTGTCCTCAAGATACAGTGG 877
D8 89145 CTTTACGCTGACATCCCGCTGACAACTCATTTCCAGAGAGCGAGTGAAGCTCACTA 89204
QY 878 TCGCGGATAGGAATCAAAAAGATACGAATGTTTGTATGATCGGTGCGGCTAAC 937
D8 89205 TTTCTGGTATCGATATACCTAAAGGCGCTGTAGTATTTGTATGATGTCGCGCAGCTAATC 89264
QY 938 GGGACCTCGAAGCTTTGAACAGCTGACGTGTTTAATTTCAATTCATCGGAAGATCTGGTA 997
D8 89265 GTGATCCATCGGTTTTTCAAAACCAATTAATTTGATCTATATCGAAGAAAATACCA 89324
QY 998 TCAAGAGCGCTTTTAGCGCGCGCGCGCTCTCGCTTTTCGGATCCCGCAATTCATACT 1057
D8 89325 CTTCTCCACAGAAAGCAATCGAAGAACATTTAGCTTTTGGTCAGGTACTCATGCT 89384
QY 1058 GTGTAGGACAGCTTTTGGCAAAAACGAAATCGAAATTTGTAGCTAATATTTGTGCTGATA 1117
D8 89385 GTGCTCGCGCTGCAATTTTCACTCAGTCAGTTGGAAGTTTCAATCAATATTTTGGATC 89444
QY 1118 AGATCGGGAATATCAGATTAGAGGAAGATTTTGTGTTATGCTGAGTCCGCTCTGTATACAC 1177

D8 89445 TCCTGATAAATTTAGCTTTTGTCTGATCACTATTAATCAAGAAACAGCGGTGTATACAC 89504
QY 1178 GCGGACCTGTTTTCACCTTCTCGTTGCGTTGA 1208
D8 89505 GAGGACCTTCTAAACTCCTTTTGTAGTTTGA 89535
RESULT 6
AAA95667
ID AAA95667 standard; DNA; 1204 BP.
XX
AC AAA95667;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating enzyme gene #2.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS Bacillus subtilis.
XX
PN WO200044886-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-JP000472.
XX
PR 29-JAN-1999; 99JP-00021707.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX
DR WPI; 2000-548827/50.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors.
XX
PS Claim 16; Page 90-94; 111pp; Japanese.
XX
CC This sequence represents the coding region for a novel Bacillus derived
CC protein having the activity of producing a hydroxylated bicyclic compound
CC or the corresponding lactone from a bicyclic compound or the
CC corresponding lactone. The protein is used for preparing hydroxylated
CC compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC inhibitors and hypocholesterolaemic agents
XX
SQ Sequence 1204 BP; 333 A; 253 C; 316 G; 302 T; 0 U; 0 Other;
Query Match 6.7%; Score 82; DB 3; Length 1204;
Best Local Similarity 45.7%; Pred. No. 5.3e-15;
Matches 329; Conservative 0; Mismatches 385; Indels 6; Gaps 1;
QY 270 AATTGTAGTGAAAGCTTTATCGGTGACGCACTGATCATCTGTCTCCATTTGATTAAACA 329
D8 289 AGTCGTGAACAAGAGCTTTTACTCCGCGCGATGAAGCAATCGGAACCGAGAATTCACGA 348
QY 330 AAATCGCAAAACTTGTAGCGCTTATCTTGAAGAGGAAAGTATCTCTCAATGA 389
D8 349 AATCAACATGAACATGATTCAAAAATTTGAGGGCGCATGTAGTTTGAACCTTTTCAACA 408
QY 390 TTTTGGAAAGAGCTTTGCGGTGTCGCTACCATGACATGCTCGGCTCGATAAAGAGA 449
D8 409 TTTTTCATACCCGCTTCGGTTATTGTATCTGAGCTGCTGGAGTGCCTTCAGCGCA 468
QY 450 CAATGAAAAAATCTCTGAGTGGCACAGCGGAGTTCGCGAGTTTATCACGAGTATCTCTCA 509
D8 469 TATGGAACAGTTTTAAAGCATGTCTGTGATCTTCTGTCTAGTACACCGAAGGATAAAGTGA 528

QY 510 ATCTCTGAA-----GCCGGGCAATTCGTTATGGTCAGCGAAGACAGCTTCCCAATA 563
 Db |||||
 529 AGAAGCTGAAAAGCCCTTTTGGAAAGCAGATAGTGTGAGGAAGAACTGGCCGCGTT 588
 QY 564 CTTGATGCCGGTCATTAAAGACGTCGCTCAATCCGGGATCAGATTTAATTTTCGATCCT 623
 Db |||||
 589 TTTTGGCGGCATCATAGAAGAAAGCGAAACAAACCGGAACAGGATATTATTCTATTTT 648
 QY 624 ATGTACTTCTGAATATGAAGGCATGCGCTGTGGCAAGGATATATCTCGCACTGATTC 683
 Db |||||
 649 AGTGAAGCGGAAGAAACAGGCGAGAAGCTGTCCGCTGAAGAGCTGATTCGTTGTGCAC 708
 QY 684 TAAATGTCTGTAGCCGCAACCGAACCGGCTGATTAAGACGCTGGCACTGATCTACCA 743
 Db |||||
 709 GCTGCTGTGGTGGCGGAAATGAACACCACTCAAACTGATTTCAAAATGCGATGTACAG 768
 QY 744 TTTTGCTCAACAATCTGACAGATGAATGATGTTTGGCTGACCGTTCTGTAGTTCCGAG 803
 Db |||||
 769 CATATTAGAAACGCCAGGGCTTACAGGAACCTGGCAGCCATCTCTGAACATGATGCTCA 828
 QY 804 AGCCATTGCGGAGACATTCGCTTATAAACCGCGGTTCAAGCTGATTCGCGGCAAGCTGTC 863
 Db |||||
 829 GGCAGTGGAGGAAGCCCTTGGTTTCAGAGCGCGCGCCCGGTTTTCAGGCGCATTCGCCAA 888
 QY 864 CCAAGATACAGTGTGCGGGTATGGAATCAAAAAGATACGATTTGTTTGTATGAT 923
 Db |||||
 889 GCGGATACGAGATCGGGGGGCACCTGATTAAGAAGGTGATGATGTTTGGCGTTGT 948
 QY 924 CGGTGGGCTAACGGGACCTGAAGCATTTGAACAGCCTGACGCTGTTAATATTATCATCG 983
 Db |||||
 949 GGCATCGGCAATCGTGATGAAGCAAAAGTTTGACAGACCGCACATGTTTGATATCGCGC 1008

RESULT 7
 ACF06114
 ID ACF06114 standard; DNA; 1188 BP.
 XX
 AC ACF06114;
 DT 30-SEP-2003 (first entry)
 DE Bacterial P450 enzyme encoding DNA SEQ ID NO:11.
 XX
 KW Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
 KW diol; alkene; chirality; thermotolerance; thermostability; gene; ds.
 XX
 OS Eubacteria.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1188
 FT /*tag= a
 FT /product= "P450 enzyme"
 XX
 PN WO2003052050-A2.
 XX
 PD 26-JUN-2003.
 XX
 PF 05-AUG-2002; 2002WO-US024910.
 XX
 PR 03-AUG-2001; 2001US-0309497P.
 XX
 PA (DIVE-) DIVERSA CORP.
 XX
 PI Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short J;
 XX
 DR WPI; 2003-541641/51.
 DR P-PSDB; ABR82124.
 XX
 PT Novel polypeptide having P450 enzyme activity and polynucleotides
 PT encoding the polypeptide, useful for catalyzing the hydrolysis of
 PT epoxides and arene oxides to their corresponding diols.
 XX
 PS Claim 18; Page 56; 365pp; English.

XX The present sequence encodes a bacterial P450 enzyme. P450 enzymes can be
 CC used to catalyse the hydrolysis of epoxides and arene oxides to their
 CC corresponding diols. P450 enzymes can also be used for hydrolysing an
 CC alkene for producing a compound of a desired chirality, and for
 CC increasing thermotolerance or thermostability of a P450 polypeptide
 XX
 SQ Sequence 1188 BP; 239 A; 371 C; 338 G; 240 T; 0 U; 0 Other;
 Query Match 6.7%; Score 81.4; DB 9; Length 1188;
 Best Local Similarity 52.2%; Pred. No. 8.1e-15;
 Matches 181; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
 QY 640 GAAGGATGCGCTGTGGACAAGGATATCTCGCACTGATTCCTTAATGTGCTGTAGCC 699
 Db |||||
 637 CATGGCCGCCCTCGACGATCAGCGCTGTGGCCAATATCTTTTCGCTGTGATTCAG 696
 QY 700 GCAACGGACCGCTGATAAGACGCTGGCACATGATCTACCATTTGCTCAACATCCT 759
 Db |||||
 697 GGCTCGGACACCGTGCCTGTCATCGGGCGGCGCAATCTATTCTGCGGAGCATCG 756
 QY 760 GAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGACCAATTCGGAGACA 819
 Db |||||
 757 GCGAGCTGGAGCGTGGCTCCGACCGCGGCTGATTCGCCGCCCTTCGCTGAGACC 816
 QY 820 TTGCTTATAAACCCCGGTTCCAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGGTC 879
 Db |||||
 817 GTGCGCTACGATCAGCGCACCAATGCTACTGGGCGGACTGCTTCCCATTCACACCGACAAA 876
 QY 880 GCGGATGGAATCAAAAAGATACGATTTTGTATGATCGGTGGGCTACCGG 939
 Db |||||
 877 TACGCAAGCCGATGAAAAAGGTCAAGCGGCTCTTTCATGATGTCGTCGGGAAACCGT 936
 QY 940 GACCCTGAAGCATTTGAACAGCCTGACGCTGTTTAATATTATTCATCGGA 986
 Db |||||
 937 GACCCGCTGGAATTCGACACACCCCGACAGCTTCATATATATACCGGA 983

RESULT 8
 AAA95666
 ID AAA95666 standard; DNA; 1191 BP.
 XX
 AC AAA95666;
 DT 14-FEB-2001 (first entry)
 DE Bacillus subtilis hydroxylating enzyme gene.
 XX
 KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
 KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia..
 XX
 OS Bacillus subtilis.
 XX
 PN WO200044886-A1.
 XX
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000WO-JP000472.
 XX
 PR 29-JAN-1999; 99JP-00021707.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
 XX
 DR WPI; 2000-548827/50.
 DR P-PSDB; AAB15501.
 XX
 PT New protein derived from Bacillus genus microorganism useful for
 PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
 PT inhibitors.
 XX
 PS Claim 14; Page 64-67; 111pp; Japanese.

QY 804 AGCCATTGGGAGACATTCCGTTTATAAACCGCGGTTTCACTGATTCGCGCGAGCTGTC 863
DB 846 GCGAGTGGAGGAGCCTTGGGTTTTCAGAGCGCGCGCGCGGTTTTCAGCGGCAATTGCCAA 905
QY 864 CCAAGATACAGTGGTGGCGGATGGAATCAAAAAGATACAGATGTTTTTGTATGAT 923
DB 906 GCGGATACGGAGATCGGGGGCACCTGATTAAAGAGGTGATGTTTTGGCGTTGT 965
QY 924 CGGTGGGCTAACCGGACCTGAAGCATTGAACAGCGCTGAGTGTGTTTAAATATTCATCG 983
DB 966 GGCATCGGCNAATCTGATGAAGCAAGTTTCACAGACCGCACATGTTTGATATCGCGG 1025

RESULT 10
AAA95669
ID AAA95669 standard; DNA; 1221 BP.
XX
AC AAA95669;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating enzyme gene #4.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; ds;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
OS
OS Bacillus subtilis.
PN WO200044886-A1.
XX
PD 03-AUG-2000.
PF 28-JAN-2000; 2000WO-JP000472.
XX
PR 29-JAN-1999; 99JP-00021707.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX WPI; 2000-548827/50.
DR P-PSDB; AAB15503.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors.
XX
PS Claim 16; Page 101-104; 11pp; Japanese.
XX
CC This sequence represents the coding region for a novel Bacillus derived
CC protein having the activity of producing a hydroxylated bicyclic compound
CC or the corresponding lactone from a bicyclic compound or the
CC corresponding lactone. The protein is used for preparing hydroxylated
CC compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC inhibitors and hypocholesterolaemic agents
XX
SQ Sequence 1221 BP; 334 A; 261 C; 322 G; 304 T; 0 U; 0 Other;

Query Match 6.4%; Score 77.2; DB 3; Length 1221;
Best Local Similarity 45.3%; Pred. No. 1.8e-13;
Matches 326; Conservative 0; Mismatches 388; Indels 6; Gaps 1;

QY 270 AATTGTAGTGAGAGCTTTATCGTGACGACCTGGATCATCTCTCCATTGATTAAACA 329
DB 306 AGTCGTGAACAAGCCCTTTACTCCGCGCGATGAAGCAATGGGAACCGAGAAATCAAGA 365
QY 330 AAATGCAAGAAATCTGTAGCGCTTATCTGAAAGAGGAAAGTATCTGTCATCAATCA 389
DB 366 ATCAAGATGACTGATTCAAAATTTTCAGGGCGCAGTGAGTTTGACCTTGTTCACGA 425
QY 390 TTTTGGAAAGAGCTTTGGCGGTGTCACGATGGACATGCTCGGCTGGATAAAGAGA 449

DB 426 TTATTATACCCGCTTCGGTTATTGTGATATCTGAGCTGCTGGAGTGCTTCAGCGCA 485
QY 450 CCATGAAAAAATCTCTGAGTGGCACAGCGGAGTTCCCGATTATATCACAGTATCTCTCA 509
DB 486 TATGGAACAGTTTAAAGCATGGTCTGATCTTCTGTGTCAGTACACCGAAGGATAAAGTGA 545
QY 510 ATCTCTGAA-----GCGCGGGACATTCGTTATGTTGTCAGCGAACAGCTTTCCCAATA 563
DB 546 AGAAGCTGAAAAGAGCTTTTGGAAAGAACGAGATAGTGTGAGGAAGAACTGGCGCGCTT 605
QY 564 CTTGATGCGGTCATTAAAGAACGTCGCGTCAATCCGGGATCAGATTTAAATTTTCATCTCT 623
DB 606 TTTTCCCGCATCATAGAGAAAGCGAAACAAACCGGAAACAGGATATATTTCTATTTT 665
QY 624 ATGTACTTCTGAATATGAAGCATCGCGTGTTCGACAAAGGATATATCTGCGCACTGATCT 683
DB 666 AGTGAAGCGGAAGAAACAGCGGAGAGCTGTCCGCTGAAGAGCTGATTCCTGTTGTCAC 725
QY 684 TAATGTCTGTAGCGCAACCGAACCGCTGATAGAGCTGGCACTGATGATCTACCA 743
DB 726 GCTGCTGCTGGTGGCGGAAATGAACCACTACAAACCTGATTTCAAAATGCGATGTTTCA 785
QY 744 TTTGCTCAACATCTCTGAGCAGATGAATGATGTTTGGCTGACCGTTCGTTAGTTCCGAG 803
DB 786 CATATTAGAACCGCAGCGGTTTACGAGAACTGGCAGCCATCTGAACTGATGCCCA 845
QY 804 AGCCATTGGGAGACATTGCGTTATAAACCGCGGTTTCAGCTGATTCGCGCGCAGCTGTC 863
DB 846 GGCAGTGGAGGAAGCCTTGGCGTTTCAGAGCGCGCGCGGTTTGGAGCGCATTCGCCAA 905
QY 864 CCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATGTTTTTGTATGAT 923
DB 906 GCGGATACGAGATCGGGGGGCACCTGATTAAAGAGGTGATACGGTTTTTGGCGTTGT 965
QY 924 CGGTGCGCTAACCGGACCTCTGAAGCATTTTGAACAGCTGACGTTTAAATATTCATCG 983
DB 966 GGCATCGGCAATCTGTGATGAAGCAAGTTTTCAGACCGCACATGTTTGATATCGCGCG 1025

RESULT 11
ABK74899
ID ABK74899 standard; DNA; 966 BP.
XX
AC ABK74899;
XX
DT 13-AUG-2002 (first entry)
XX
DE Bacillus licheniformis genomic sequence tag (GST) #2190.
XX
KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
XX physiological provocation; ds.
OS Bacillus licheniformis.
XX
PN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US031437.
XX
PR 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX
PA (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX
PI Berka R, Clausen IG;
XX WPI; 2002-416684/44.
XX
PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus

cells, by using substrate containing *Bacillus* genomic sequenced tag array.

Claim 4; SEQ ID NO 2190; 200pp; English.

The invention describes a method of monitoring differential expression of genes in a first *Bacillus* cell relative to expression of the genes in other *Bacillus* cells, comprising hybridising labelled nucleic acid probes isolated from *Bacillus* cells to a substrate containing array of *Bacillus* genomic sequenced tags (GST), examining the array, and determining relative gene expression by an observed hybridisation reporter signal of a spot in the array. The method is useful for measuring the expression of genes in a first *Bacillus* cell relative to expression of the same genes in one or more second *Bacillus* cells. The method is useful for monitoring global expression of several genes from a *Bacillus* cell, discovering new genes, identifying possible functions of unknown open reading frames and monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which *Bacillus* cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at pub.wipo.int/pub/published pct sequences

Sequence 966 BP; 252 A; 229 C; 269 G; 216 T; 0 U; 0 Other;

Query Match 6.0%; Score 72.4; DB 6; Length 966;

Query Match: 0.0%; Score: 72.4; DS: 0;
Best Local Similarity 45.3%; Pred. No. 5.6e-12;

Sequence similarity	Matches	Conservative	Mismatches	Indels	Gaps
100% identical	345	0	411	6	2

228 GCTGGCCCAAATGCATGGAAAGAACACTCTGCCAAAGAGAAATTGTAGTGAGAGCTT 287

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

288 TATCGGTGACGCACCTGGATCATCTGTCTCCATTGATTAAACAATAATGCAGAAACTTGTT 347

77 TACACCGCGCTCATGAAGGAGTGGGAGCCGCGCATTCGCGAACTGACGAATCAATTGCT 136

[illegible]

137 CGCTGATGTCGCGGAGAGATCGACCTTGACAGGATTTTCATATCCGTTGCC 196

408 GGTGTGCGTCACGATGGACATGCTCGGGCTGGATAAAGAGACCATTGAAAAATCTCTGA 467

468 GTGGCACAGCGGAGTTGCCGATTTATCACGAG-TATCTCTCAATCTCTGAAGCGCGGG 526

257 ATGGTCCGACCTGTTGGTCAGCCTGCCGAAGAGCGATCGGCCGAAGATGTGAACGAGTG 316

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

317 GAAAAACATCAGGGACCAAGCGAAGAGCTGACCGCATTCTTTGAAAAGATGATTGA 376

582 AGAACGTCGGTCAATCCGGATCAGATTAAATTCGATCCTATGTACTTCTGAATGA 641

123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

642 AGGCATGGCGCTGTCGGACAAGGATATACTCGCACTGATTCTTAATGTGCTGTAGCCGC 701

Query Match 5.6%; Score 67.8; DB 4; Length 110000;

Query Match	3.0%	Score 07.0; DS 4; Bengali 110000
Best Local Similarity	45.1%	Pred. No. 2.2e-09;

Matches	292;	Conservative	0;	Mismatches	352;	Indels	3;	Gaps	1;
----------------	-------------	---------------------	-----------	-------------------	-------------	---------------	-----------	-------------	-----------

337 GAAAACTTGTAGCGCCTATCTTGAAAGAGGGAAGTGATCTCGTCAATGATTTGGA 396

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

XX Bacillus licheniformis genomic sequence tag (GST) #2182.
DE
XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
XX Bacillus licheniformis.
XX
XX WO200229113-A2.
XX
XX 11-APR-2002.
XX
XX 05-OCT-2001; 2001WO-US031437.
XX
XX 06-OCT-2000; 2000US-00680598.
XX 27-MAR-2001; 2001US-0279526P.
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
XX (NOVO) NOVOZYMES AS.
XX
XX Berka R, Clausen IG;
XX
XX WPI; 2002-416684/44.
XX
XX Monitoring differential expression of several genes in first Bacillus
XX cell relative to expression of same genes in one or more second Bacillus
XX cells, by using substrate containing Bacillus genomic sequenced tag
XX array.
XX
XX Claim 4; SEQ ID NO 2182; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX relative gene expression by an observed hybridisation reporter signal of
XX a spot in the array. The method is useful for measuring the expression of
XX genes in a first Bacillus cell relative to expression of the same genes
XX in one or more second Bacillus cells. The method is useful for monitoring
XX global expression of several genes from a Bacillus cell, discovering new
XX genes, identifying possible functions of unknown open reading frames and
XX monitoring gene copy number variation and stability. Monitoring changes
XX in expression of genes may be used to provide a representation of the way
XX in which Bacillus cells adapt to changes in culture conditions,
XX environmental stress or other physiological provocation. Extensive follow
XX -up characterisation is unnecessary, when one spot on an array equals one
XX gene or one open reading frame, since sequence information is available.
XX This sequence represents a genomic sequence tag (GST) used in the method
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1213 BP; 340 A; 286 C; 287 G; 300 T; 0 U; 0 Other;
XX
Query Match 5.3%; Score 64.8; DB 6; Length 1213;
Best Local Similarity 47.0%; Pred. No. 1.7e-09;
Matches 201; Conservative 0; Mismatches 227; Indels 0; Gaps 0;
XX
QY 560 AATACTTGTATGCGGTTCATTAAAGAACGTCGCGTCAATCGGGATCAGATTTAAATTGCA 619
DB 589 ATATTTTAAACATTTTCAACGACGCGCAAAAAGAGCGCAAAAGACGACCTGATTTCCG 648
QY 620 TCCTATGTACTTTCTGAATATGAAGGATGCGGTGTCGGAACAAGGATATATCGCACTGA 679
DB 649 TTTTACTGCGGGCGGAAGTTGACGGCAATCGCTGACAGAAGAAGAACTGCTTCATTTT 708
QY 680 TTCCTTAATGTGTTGTAGCGCAACGACCGGCTGATAGACGCTGGCGCATGATGATCT 739
DB 709 GCATCATCTTTTGGTCGAGGCAATGAGACGCAACCAACTTGATCGCAAAACGCGTCC 768
QY 740 ACCATTGTCTCAACAATCCTGACAGATGAATGATGTTTTGGCTGACCGTTCGTTAGTTC 799

RESULT 15

AAI99682_08/c

Continuation (9 of 45) of AAI99682 from base 800001 (Mycobacterium tuberculosis strain H:
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
----	---------------	-------	-----

WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	441529

Query Match

Best Local Similarity 5.1%;

Score 61.8; DB 4; Length 110000;
Pred. No. 1.9e-07;
46.8%;

Matches 195; Conservative		0;	Mismatches 222;	Indels	0;	Gaps	0;
Qy	608	ATTAAATTTTCGATCTTACTTCTGAATATGAAGGCATGGCGCTGTTCGGACAAGGATA	667				
Db	59445	ATCTGACGTCGGCACTGTCTGGCAGCCGAGCTCGACGGCGACCGGCTTTCCGACCAGGAAA	59386				
Qy	668	TACTCGCACTGATCTTAATGTGCTGTAGCCGCAACGGACCGGCTGATAGACCGCTGG	727				
Db	59385	TCATGGCGTTCCTGTCTTCATNGGTGATCGCCGGCAACGAGACCAACCAAGCTACTGG	59326				
Qy	728	CACTGATGATCTACCATTTGCTCAACAATCCTGAGCAGATGAATGATGTTTTGGCTGACC	787				
Db	59325	CCAATGCCGTCTACTGGGCGGCCCAACCCCTGGCCAGCTGGCCCGGTATTGCCCGACC	59266				
Qy	788	GTTTCGTTAGTTCGAGAGCCATTCGGGAGACATTCGGTTATAAACCGCCGGTTCAGCTGA	847				
Db	59265	ACTCCCGGATTCGATGTGGGTGGAGGAAACCTCGCTACGACAGTCCAGCCAGATTC	59206				
Qy	848	TTCCCGGGCAGCTGTCCCAAGATACAGTGTGGCGGTATGGAAATCAAAAAAGATACGA	907				
Db	59205	TGGCCCGCACCGTCCGCGCACGATCTCACGTTGTACGACACCAACGATCCCGAGGGTGAG	59146				
Qy	908	TTGTTTTTTGTATGATCGGTGGGCTAACCGGACCCCTGAAGCATTTGAACAGCCTGACG	967				
Db	59145	TGTTGCTGTCTACCGGATCGGCCAACGTAACGACCGGGTGTTCGAGACCCCGACG	59086				
Qy	968	TGTTTAATATTCATCGGGAAGATCTTTGGTATCAAGAGCGCTTTTAGCGCGCGCCCC	1024				
Db	59085	ACTATCGCATCGGCCGCGAAATCGGCTGCAAACTAGTCTAGTTTCGCGAGCGGTGCC	59029				

Search completed: December 10, 2004, 13:46:19
Job time : 681.942 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 16:12:37 ; Search time 721.715 Seconds
(without alignments)
9281.958 Million cell updates/sec

Title: US-10-627-124-1
Perfect score: 1215
Sequence: 1 atgagccaatcgattaaatt.....tcgttgcttgacggggca 1215

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq.*
- 20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
- 21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1215	100.0	1215	16	US-10-627-124-1
2	617.2	50.8	1208	9	US-09-974-300-2181
3	113.6	9.3	2731748	17	US-10-297-465A-1
4	89.4	7.4	2731748	17	US-10-297-465A-1
5	81.4	6.7	1188	15	US-10-214-446-11
6	72.4	6.0	966	9	US-09-974-300-2190
7	67.6	5.6	1215	15	US-10-156-761-1
8	67.6	5.6	9025608	15	US-10-156-761-1
9	64.8	5.3	1213	9	US-09-974-300-2182
10	55.6	4.6	1200	16	US-10-282-122A-25888
11	55.6	4.6	84428	15	US-10-229-148B-1
12	55.2	4.5	1233	15	US-10-205-032-7

c	13	55.2	4.5	60196	15	US-10-205-032-1	Sequence 1, Appli
	14	55	4.5	1155	9	US-09-974-300-2185	Sequence 2185, Ap
	15	53.6	4.4	1215	15	US-10-214-446-49	Sequence 49, Appl
	16	53.4	4.4	1248	15	US-10-156-761-7447	Sequence 7447, Ap
c	17	50	4.1	17596	18	US-10-611-442-2	Sequence 2, Appli
	18	49.6	4.1	1191	15	US-10-214-446-13	Sequence 13, Appl
	19	48.2	4.0	9024	18	US-10-611-442-1	Sequence 1, Appli
	20	46.6	3.8	1290	9	US-09-738-626-617	Sequence 617, App
c	21	46.6	3.8	1407	17	US-10-781-014-779	Sequence 779, App
	22	46.6	3.8	3309400	9	US-09-738-626-1	Sequence 1, Appli
	23	46.2	3.8	1248	15	US-10-214-446-55	Sequence 55, Appl
c	24	44.8	3.7	82993	15	US-10-080-170-645	Sequence 645, App
	25	44.8	3.7	82993	17	US-10-080-170-645	Sequence 645, App
c	26	44.6	3.7	82993	18	US-10-468-356-645	Sequence 39, Appl
	27	44.6	3.7	1215	15	US-10-214-446-39	Sequence 94, Appl
	28	43.6	3.6	1293	14	US-10-145-415-94	Sequence 21, Appl
	29	43	3.5	1209	9	US-09-861-289-21	Sequence 21, Appl
	30	43	3.5	1209	9	US-09-860-846-21	Sequence 21, Appl
	31	43	3.5	1209	10	US-09-988-384B-21	Sequence 21, Appl
	32	43	3.5	1209	15	US-09-836-821-21	Sequence 21, Appl
	33	43	3.5	1209	15	US-10-271-889-21	Sequence 21, Appl
	34	43	3.5	1209	17	US-10-398-605-21	Sequence 3, Appli
	35	43	3.5	12441	10	US-09-988-384B-3	Sequence 3, Appli
	36	43	3.5	13613	9	US-09-861-289-3	Sequence 3, Appli
	37	43	3.5	13613	9	US-09-860-846-3	Sequence 3, Appli
	38	43	3.5	13613	10	US-09-836-821-3	Sequence 3, Appli
	39	43	3.5	13613	15	US-10-271-889-46	Sequence 46, Appl
	40	43	3.5	13613	17	US-10-398-605-3	Sequence 3, Appli
	41	43	3.5	38506	10	US-09-793-708-19	Sequence 19, Appl
	42	43	3.5	38506	15	US-10-201-385-1	Sequence 1, Appli
	43	43	3.5	38506	15	US-10-160-539-19	Sequence 19, Appl
	44	42.6	3.5	1179	15	US-10-156-761-1975	Sequence 1975, Ap
c	45	42.6	3.5	9025608	15	US-10-156-761-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-10-627-124-1
; Sequence 1, Application US/10627124
; Publication No. US20040096944A1
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient
; TITLE OF INVENTION: Mutants Of Bacillus Cells
; FILE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; PRIOR FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-627-124-1

Query Match	100.0%	Score 1215;	DB 16;	Length 1215;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1215;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGAGCCAAATCGATTAAATTTGTTTGTCTGTCATCAATTTCAAAACCAATCCATAT	60	
DB	1	ATGAGCCAAATCGATTAAATTTGTTTGTCTGTCATCAATTTCAAAACCAATCCATAT	60	
QY	61	GCTTATTTTTCACAACTCGCGGAGGAGATCCGGTTTCATTTATGAAGTCGATAGACAGT	120	
DB	61	GCTTATTTTTCACAACTCGCGGAGGAGATCCGGTTTCATTTATGAAGTCGATAGACAGT	120	
QY	121	TATTTTATCAGCGGCTATCATCATGTCCTTATATCTTTCAGATCCGATATCTTCACG	180	
DB	121	TATTTTATCAGCGGCTATCATCATGTCCTTATATCTTTCAGATCCGATATCTTCACG	180	

Applicant

```
Db 121 TATTTTATCAGCGCTATCATGATGTCGCTATATCTCTTACGATCCGGATATCTTCACG 180
Qy 181 ACGAATCACTTTGAGCGTCCGACCAAGCAGTCATCGGAGCGCTGCTGCGCCCAATG 240
Db 181 ACGAATCACTTTGAGCGTCCGACCAAGCAGTCATCGGAGCGCTGCTGCGCCCAATG 240
Qy 241 CATGAAAAGAACACTCTGCCAAAAGAAATGTTAGTGAGAGCTTTTATCCGTCACGCA 300
Db 241 CATGAAAAGAACACTCTGCCAAAAGAAATGTTAGTGAGAGCTTTTATCCGTCACGCA 300
Qy 301 CTGATCATCTCTCTCCATGTATTAACAAAATGACAGAAAATGTTAGTCGCTTATCTT 360
Db 301 CTGATCATCTCTCTCCATGTATTAACAAAATGACAGAAAATGTTAGTCGCTTATCTT 360
Qy 361 GAAAGGGGAAAGTATCTCTCATGATTTTGGAAAGACCTTTGCGGTGTCGCTCAG 420
Db 361 GAAAGGGGAAAGTATCTCTCATGATTTTGGAAAGACCTTTGCGGTGTCGCTCAG 420
Qy 421 ATGGACATCTCGGCTGGATAAAGAGACCATGAAAAATCTCTGAGTGGCACAGCGGA 480
Db 421 ATGGACATCTCGGCTGGATAAAGAGACCATGAAAAATCTCTGAGTGGCACAGCGGA 480
Qy 481 GTTGGCGATTTTATCACAGATATCTCTCAATCTCTGAAGCGCGGCACATTCGTTATGG 540
Db 481 GTTGGCGATTTTATCACAGATATCTCTCAATCTCTGAAGCGCGGCACATTCGTTATGG 540
Qy 541 TCGAGGAAACAGCTTCCCAATCTGATGCGCGTCAATTAAGAGAGCTGCGTCAATCCG 600
Db 541 TCGAGGAAACAGCTTCCCAATCTGATGCGCGTCAATTAAGAGAGCTGCGTCAATCCG 600
Qy 601 CGATCAGATTTAAATTCGATCTCTATGATCTCTGAATATGAAGGATGCGCTGTCGAC 660
Db 601 CGATCAGATTTAAATTCGATCTCTATGATCTCTGAATATGAAGGATGCGCTGTCGAC 660
Qy 661 AAGATATATCTCGACTGATTTCTTAATGTCTTTAGCCGCAACCGCTGATAG 720
Db 661 AAGATATATCTCGACTGATTTCTTAATGTCTTTAGCCGCAACCGCTGATAG 720
Qy 721 ACGCTGGCATGATGATCTACATTTGCTCAACATCTCGAGCAGATGATGTTTG 780
Db 721 ACGCTGGCATGATGATCTACATTTGCTCAACATCTCGAGCAGATGATGTTTG 780
Qy 781 GCTGACCGTTCTGTTAGTCCGAGAGCAATGCGGAGACATGCGTTATATAACCGCGTT 840
Db 781 GCTGACCGTTCTGTTAGTCCGAGAGCAATGCGGAGACATGCGTTATATAACCGCGTT 840
Qy 841 CAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAA 900
Db 841 CAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAA 900
Qy 901 GATACGATTTGTTTGTATGATCGGTGCGCTAACCGGACCTGGAAGCATTTGAACAG 960
Db 901 GATACGATTTGTTTGTATGATCGGTGCGCTAACCGGACCTGGAAGCATTTGAACAG 960
Qy 961 CTGACGCTGTTTAAATTCATCGGGAAGATCTTGATCAAGAGCGCTTTTAGCGCGCC 1020
Db 961 CTGACGCTGTTTAAATTCATCGGGAAGATCTTGATCAAGAGCGCTTTTAGCGCGCC 1020
Qy 1021 GCCCGGCATCTCGCTTTCCGATCCGGCATTCATACTGTGTAGGAGCAGCTTTTGCCAA 1080
Db 1021 GCCCGGCATCTCGCTTTCCGATCCGGCATTCATACTGTGTAGGAGCAGCTTTTGCCAA 1080
Qy 1081 AACGAATCGAATTTGATGATGATGCTGCGGATTAAGATCGGGAATATCAGATTAGAG 1140
Db 1081 AACGAATCGAATTTGATGATGATGCTGCGGATTAAGATCGGGAATATCAGATTAGAG 1140
Qy 1141 GAAAGATTTTGTATGCTGAGTCCGCTGTTATACACGCGGACCTGTTCACTTCTCGTT 1200
Db 1141 GAAAGATTTTGTATGCTGAGTCCGCTGTTATACACGCGGACCTGTTCACTTCTCGTT 1200
Qy 1201 GCGTTTGACGGGCA 1215
Db 1201 GCGTTTGACGGGCA 1215
```

RESULT 2

```
US-09-974-300-2181
; Sequence 2181, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2181
; LENGTH: 1208
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2181
```

```
Query Match 50.8%; Score 617.2; DB 9; Length 1208;
Best Local Similarity 70.5%; Pred. No. 1.4e-185;
Matches 853; Conservative 0; Mismatches 353; Indels 4; Gaps 2;

Qy 1 ATGAGCCCAATCGATTAAATTTGTTAGTGTGCTTTCTGATCAATTTCAAAACAATCCCATAT 60
Db 1 ATGATCAATCGCTAAAACAATTCAGCGTGTGTGTCAGACATATATCAGAAAATCCGTAT 60
Qy 61 GCTTATTTTCAAACTCGGGAGGAAGATCCGGTTCTATTATGAAGATCGATAGACAT 120
Db 61 CAGTATTTTCTATACCTTTCCGGAATCTGATCCCGTCCATTTATGAAGAATCGCTGGACAG 120
Qy 121 TATTTTATCAGCGCTATCATGATGTCGCTATATCTTTCAGCATCCGATATCTTCACG 180
Db 121 TATTTTATCAGCGCTATCAAGATGTGCGCGCTGCTCAGAAATCAGGACGCTTTTACA 180
Qy 181 ACGAAAT--CACTTCTTGAGCGTGCAGAACCAAGTATGCGAGCGCTGCTGCGGCCAAA 238
Db 181 ACGATTTGCTGTGCTGCCAAAACGGCGCGAGCCGCTCATGCGCGGACCTGTGCTGCCCAA 240
Qy 239 TGCATGGAAGAAACACCTCTGCAAAAGAAATTTGTAGTAGAAGACTTTATCGGTGACG 298
Db 241 TGAAGGCAAGAGACACACGCGCAAAAGAGAAATCGTTCTGCGCGCTTTATCGGGGAAT 300
Qy 299 CACTGATCATCTGCTCCATTTGATTAAACAAATGCAGAAAATTTGTTAGCGCTTATC 358
Db 301 CCGTTGATCACCTCACACCGCTCATCAAGAAAATGCCAAAAGGCT--TTTGGCCCCGCAC 358
Qy 359 TTGAAAGAGGAAAAAGTATCTCGTCAATGATTTTGGAAAAGACGTTTGGCGGTGTGCGTCA 418
Db 359 GTGGAGAGGGCGGATCGACCTTGTCAATGATTTGCGCAAAACATTTGCGCGTTTGGGTGA 418
Qy 419 CGATGGACATGCTCGGGCTGGATAAAGAGACCATGAAAAAATCTCTAGTGGCAACGCG 478
Db 419 CGATGGACATTTTAGTTTGGACAAAGACGACCAACCAACCGGTGAGAAAATCGGCACAGCG 478
Qy 479 GAGTTGCGGATTTTATCAGAGTATCTCTCAATCTCTGAAAGCGGGGACATTCCTTAT 538
Db 479 GCGTCCGCGATTTTATTAACCACTTTGAAATCAGCGCGCTGAGGACCGGGAGCATTTCCCTCA 538
Qy 539 GGTGACAGGAAACAGCTTTTCCCAATATCTTGATCGCGGTCAATTAAGAAACGTCGCGTCAATC 598
Db 539 AATGAGTGAACAGCTCGCTGAGTATTTGAATCCGATTTATCGAGGAAAGCGCGCAAAATC 598
Qy 599 CCGGATCAAGATTTAATTTTCATCTTATGATCTCTTGAATATGAAGCATGCGCTCTCGG 658
Db 599 CCGGACATGATTTAATATCTCATCTTTGCACTTCCGAATACGAGGAGTGGCGATGCTCG 658
```

QY 659 ACAAGATATACCTCGCACTGATCTTAAATGTGTCTGTAGCCGCAACGGACCGGCTGATA 718
DB 659 ACCGGACATACCGCGCTGATCTCAATATATTTGCTTCCGCGCACAGAACCGCGGACA 718
QY 719 AGACGCTGCACTGATGATCTACCATTTCTCTCAACAATCTGAGCAGATGAATGATTT 778
DB 719 AGACGCTGCACTGATGATTTATTTATTTGCTTCTCATCTCCGATCAGATGAATGATGTC 778
QY 779 TGGCTGACCGTTGCTTTAGTTTCCGAGAGCAATTTGCGGAGACATTTGCCGTTTATAACCGCCGG 838
DB 779 TGGAGACCGCAGCTTTCTCCCTCAGCAATCGCAGAGAGCTGCGCTACAGCCGCGG 838
QY 839 TTCAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAA 898
DB 839 TGCAGCTCATCCGCGCCAGCTTTTCAAAAGACCGGAGATTTGCGGAGTCGAGCTAAAAG 898
QY 899 AGATACGATTTGTTTGTATGATCGGTGCGGTACCGGACCGGACCTGAAGCATTTGAAC 958
DB 899 AAGGACGAGCTGATTTTTCATGATAGCGCGGCAAAATCGCGATCCTGAAGCGTTTCGAG 958
QY 959 AGCTGACGCTGTTTAAATTTTCAATCGGGAAGATCTTGGTATCAAGACGCTTTTAGCGGCG 1018
DB 959 ATCCGACAGTTCAACATTTACCGCAGGACCTGGAAGTCAAAAGCGATTCAGCGGCG 1018
QY 1019 CCGCCCGGATCTCGCTTTCCGATCCGCGCATTAATCACTGTGTAGGAGCAGCTTTTGCCA 1078
DB 1019 CAGCCAGGATCTCGCATTCGGTTTCAAGCGCTCCACAATCTGTGTAGGAGCGGTTTGCGA 1078
QY 1079 AAACGGAATCGAATTTAGTAAATTTGTCTGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1138
DB 1079 AGACCGAATGAGCTTTGCGCAATATCTGCTTGTATGAGTGAAGTGAAGTGAAGTGAAGTGAAG 1138
QY 1139 AGGAGATTTTGTATGCTGATCGGTCTGTATACAGCGGACCTGTTTCACTTCTCG 1198
DB 1139 AGGAGATTTTATACCGCGAGACAGGCTTTACACAGCGGACCGCTCTCGCTCAACA 1198
QY 1199 TTGCGTTTGA 1208
DB 1199 TCCGTTTGA 1208

RESULT 3

US-10-297-465A-1/c

; Sequence 1, Application US/10297465A

; Publication No. US20040142413A1

; GENERAL INFORMATION:

; APPLICANT: Simpson, Andrew

; APPLICANT: Reinach, Fernando

; APPLICANT: Setubal, Joao

; APPLICANT: Mediane, Joao

; APPLICANT: Arruda, Paulo

; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof

; FILE REFERENCE: FAPESP 202 US (10213376)

; CURRENT APPLICATION NUMBER: US/10/297,465A

; CURRENT FILING DATE: 2001-06-07

; PRIOR APPLICATION NUMBER: PCT/IB01/01618

; PRIOR FILING DATE: 2001-06-07

; PRIOR APPLICATION NUMBER: 60/209,906

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1

; LENGTH: 2731748

; TYPE: DNA

; ORGANISM: Xylella fastidiosa

US-10-297-465A-1

Query Match 9.3%; Score 113.6; DB 17; Length 2731748;

Best Local Similarity 47.5%; Pred.No. 1.1e-22;

Matches 338; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

QY 272 TTGTAGTGAAGCTTTATTCGGTGAACGCTGATCTGTCTCCATTTAAACAAA 331

DB 388822 TGGTAATGCAACGCTTCAACCGGCGAGAGATGAATCAATGCGTCCCTTGGGATCGATA 388763
QY 332 ATGCGAGAAATCTTGTAGCGCTTATCTTGAAGAGGAGGAAAAGTGTATCTCGTCAATGATT 391
DB 388762 CGCGCGACCAATTTGATCGACAATTTGAACAGAACCATCCGTTGATCTTGTGCTGAAT 388703
QY 392 TTGGAAGAACGTTTGGGCTGCGTCAAGATGGAACATGCTCGGGCTGGATATAAAGAGACC 451
DB 388702 TTGCTCTTCCCTTCCCATGCGAGATCAATTTGCAAAATGATGATGATGATGATGATG 388643
QY 452 ATGAAATAATCTCTGAGTGGCACAGCGGAGTTGCGGATTTTATCACAGAGTCTCTCAAT 511
DB 388642 CCGTCACCTCGGTATAGCGGTGAGCAAAATTTGCCAAGGATTTTCCGACCTTCCCAATGT 388583
QY 512 CTCTTGAACCGCGGCGACATTTGTTATGTTGTCAGAGCAACAGCTTTTCCCAATACTTGCATGC 571
DB 388582 CGCGCGAGTAATTTAGTACAGCAAGCACTGCTTCAAGAGCTTGGCAATACTTTACGA 388523
QY 572 CCGTCATTAAGAACGTCGCGTCAATTCGGGATCAGATTTAATTTTCGATCCTTATGACTT 631
DB 388522 AGTTGATGAGCTACGCGCACCATCTCTGAACTGACTTAATTTCTATGTTTCTCGGTG 388463
QY 632 CTGAATATGAAGCATGCGCTGTGCGCAAGATATCTCGCACTGATTTCTTAATGTC 691
DB 388462 CTGAGGAAGACGCGGAGAACTTAACCATGATGAATCTCTCAATGATGATTTATGTTAT 388403
QY 692 TGTAGCCGCAACGGAACCGGCTGATGAACGCTGCACTGATGATCTACCAATTTGCTCA 751
DB 388402 TAAATGCGGTTTACGAAACCAATCCATATGATTTGGCAATGATTTGCTTGCATC 388343
QY 752 ACAATCTGAGCAGATGAATGATTTTGGCTGACGCTTGTGTTAGTTCGAGAGCATTG 811
DB 388342 GCCATCCAGAGCAACTCGCACTCTCAAGAGTGTCTGCTGATGCGCAAGCGGTAT 388283
QY 812 CGGAGACATGCGTTTAAACCGCGGTTTCACTGATTTCCGGGACGCTCCCAAGATA 871
DB 388282 CGGAATGCTTACGTTACGACGCGGTGCTACAGTTCAGATGCGCGCGGATGATGATA 388223
QY 872 CAGTGTGCGCGGTATGGAATCAAAAAGATACGATTTGTTTGTATGATCGGTGCGG 931
DB 388222 TAGAGTTGAAGCGGAGTTGGTCCCTCGTGGCACAGTAGTATTTTGTGTTGCTG 388163
QY 932 CTAAACCGGACCTGAAGCATTTGAACGCTGAGCTGTTTAAATATTCATCG 983
DB 388162 CCAACCGTATCCGCGCAATTCATCTATCCGACCGAGCTGATTAATCTCG 388111

RESULT 4

US-10-297-465A-1

; Sequence 1, Application US/10297465A

; Publication No. US20040142413A1

; GENERAL INFORMATION:

; APPLICANT: Simpson, Andrew

; APPLICANT: Reinach, Fernando

; APPLICANT: Setubal, Joao

; APPLICANT: Mediane, Joao

; APPLICANT: Arruda, Paulo

; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof

; FILE REFERENCE: FAPESP 202 US (10213376)

; CURRENT APPLICATION NUMBER: US/10/297,465A

; CURRENT FILING DATE: 2001-06-07

; PRIOR APPLICATION NUMBER: PCT/IB01/01618

; PRIOR FILING DATE: 2001-06-07

; PRIOR APPLICATION NUMBER: 60/209,906

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1

; LENGTH: 2731748

; TYPE: DNA

; ORGANISM: Xylella fastidiosa

US-10-297-465A-1

```
Query Match          7.4%; Score 89.4; DB 17; Length 2731748;
Best Local Similarity 46.4%; Pred. No. 6e-15;
Matches 291; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

QY 376 GATCTCGTCAATGATTTTGAAGAAGAGCTTTGCGGTGTCGCTCAGCATGACATGCTCGGG 435
    |||||
DB 363766 GATCTTGTACTGAATTTTCCCTTCCCATGCGGATCATTTGACAGGATGATGGAT 363825

QY 436 CTGGATAAAGAGACATGAAAAAATCTCTAGTGGCACAGCGGAGTTGCCGATTTTATC 495
    |||||
DB 363826 GTGGATATCAGCGACGCTATCAGTCTAAGTGTGGCGGTGAGCAACATTTGCCAAGGTATTG 363885

QY 496 ACAGTATCTCTCAATCTCTGAAGCGCGGGCACATTCGTTATGTGTGACGCAACAGCTT 555
    |||||
DB 363886 GACCTTGCGCCGATGTCGCCCGATGAATTTGGTGCACGCAAGTGTCTCTACGAAGAATTA 363945

QY 556 TCCCAATACTTGATGCGCGTCAATTAAGAAACGTTCGGTCAATCCGGGATCAGATTTAAT 615
    |||||
DB 363946 GCGCATTAATCACAAGGTTGATCGAGCTGCGCGCGCGCAGCTTGGAACTGACTTGAT 364005

QY 616 TCGATCTATGACTTCTGAATATGAAGCATGGCGCTGTCGCAAGGATATATCTCGCA 675
    |||||
DB 364006 TCCATGTTGCTGCTGCTGAGGAGGAGGCCAGAAACTGACGCATGATGAAATCGTGTC 364065

QY 676 CTGATCTTAATGTCTGTTAGCCGCAACGGACCGGCTGATAAGACGCTGCACCTGATG 735
    |||||
DB 364066 AATGTGATTTTGTGTTATTTGGTGGCTACGAGACACATCCAAATGATTTGGCAATGCA 364125

QY 736 ATCTACCAATTTGCTCAACAATCTGTAGCAGATGAATGATTTTGGCTGACCGTTGCTTA 795
    |||||
DB 364126 TTGATTTGCTTGATCGTCAATCGAAGCAGCTTGACAGACTCAAGANGATTTGTCGCTG 364185

QY 796 GTTCCGAGAGCCATTCGGGAGACATTCGCTTATAAACCCCGGTTTCAGCTGATTCGCGG 855
    |||||
DB 364186 ATGCCGCAAGCATATTGGAATGCTTTACGTTACGATGGATCGTCAGTTCACGATGCGC 364245

QY 856 CAGCTGTCCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGATACGATTTGTTTT 915
    |||||
DB 364246 GCGGCGATGGATGATGTGAGCATTTGAAGGTGATGTAGTACCTCGTGGCAGGATTTGTTT 364305

QY 916 TGTATGATCGGTGCGCTCAACCGGACCTGAAGCATTTTGAACAGCTTCAGCTGTTTAA 975
    |||||
DB 364306 TTGATGCTTTGCTGCTGAACCGTATCGGCGCAATTCAGTATCCGATCATCTGGAC 364365

QY 976 ATTATCTCGGGAAGATCTTGGTATCAAG 1002
    |||||
DB 364366 ATCACACGGAAGCAAGAGACGCTCCAG 364392
```

```
RESULT 5
US-10-214-446-11
; Sequence 11, Application US/1021446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: THEM AND METHODS OF MAKING AND USING THEM
; CURRENT FILING DATE: 2002-08-05
; PRIOR FILING DATE: 2002-08-05
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1188
; TYPE: DNA
```

```
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-214-446-11

Query Match          6.7%; Score 81.4; DB 15; Length 1188;
Best Local Similarity 52.2%; Pred. No. 1.6e-14;
Matches 181; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 640 GAAGCATGGCGCTTCGCAACAAAGGATATATCTCGCACTGATTTCTTAATGTGCTTTAGCC 699
    |||||
DB 637 GATGCGCGCCCTCGACGATCAGGCGCTGTGCGCAATATCTTTTCGCTGTCGATTAGC 696

QY 700 GCAACGGAAACGGCTGATAAGACGCTGACACTGATGATCTACCAATTTGCTCAACATCTT 759
    |||||
DB 697 GGCTCGGACACCGTGCCCTGTTCATCGGCGGGGCAATCTATATCTGCGGAGCATCCG 756

QY 760 GAGCAGATGAATGATGTTTGGCTGACCGTTTCGTTAGTTCGAGAGACCAATTCGCGAGACA 819
    |||||
DB 757 GCGCAGCTGGAGCGGTGCGCTCCGACCGCGCTGATTCGCGCGCTTCGCTGAGACC 816

QY 820 TTGCGTTATAAACCGCCGCTTCAGCTGATTCGCGCGGAGCTGTCCCAAGATACAGTGTC 879
    |||||
DB 817 GTGCGCTACGATCAGCCGCAACCAATCTACTGCGCGCGACTGCTTGCCATTGACACCCGACAA 876

QY 880 GCGGCTATGGAATCAAAAAGATACGATTTGTTTTCGATGATCGGTGCGCTAACCGG 939
    |||||
DB 877 TACGCAAGCGCGATGAAAAAGGTCAAGCGGTCCTGTTTCATGATGCTGCGTCCGCAACCGT 936

QY 940 GACCCCTGAAGCATTTGAACAGCCTGACGCTGTTTAATATTCATCAGGGA 986
    |||||
DB 937 GACCCGCTGGAATTCGAACACCCCGACAGTTCAATATATACCGGA 983
```

```
RESULT 6
US-09-974-300-2190
; Sequence 2190, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2190
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2190
```

```
Query Match          6.0%; Score 72.4; DB 9; Length 966;
Best Local Similarity 45.3%; Pred. No. 1.1e-11;
Matches 345; Conservative 0; Mismatches 411; Indels 6; Gaps 2;

QY 228 GCTGGCCCAATGATGGAAGAAGACACTCTCCCAAAAGAAAGAAATTTAGTGAGAGCTT 287
    |||||
DB 17 GATGATCAGCATGGATTCGCGGAAAACACACGAGAAATCAGATCGATCGTCAACAGGCTTT 76

QY 288 TATCGGTGACGCACTGGGATCATCTGCTCCATTGATTAACAAAATGACAGAAAATTTGTT 347
    |||||
DB 77 TACACGCGCGTCATGAAGAGTGGAGCGCGGCAATTCGGAACACTGACGAATCAATGCT 136

QY 348 AGCGCTTATCTTGAAGAGGAAAAGTGAATCTCGTCAATGATTTTGGAAAGAGCTTGC 407
    |||||
DB 137 CGCTGATGTCCGCGCAGGGAAGAGATCGACCTTGTACAGGATTTTTCATATCCGTTGCC 196
```


QY 408 GGTGTCGTCACGATGACATGCTCGGGCTGGATAAAGAGACCATGAAAAATCTCTGA 467
DB |||||
197 CGTAATCGTCATTTGCGAATGCTCGGCGTCTCTTTGGTGATAAGCATCATTTTCAAGA 256
QY 468 GTGGCACAGGGGTTGCGGATTTATACGAG-TATCTCTCAATCTCTGAGCGCGG 526
DB |||||
257 ATGGTCCGACCTGTTGGTCAGCTCGCGAAGAGCATCGCGCGAAGATGTGAAAGAGTG 316
QY 527 CACATTCGTTATGG-----TGCAAGCAACAGCTTTCCCAATCTTGATGCGCGTCAATAA 581
DB |||||
317 GAAAACATCAGGACCAAGCGGCAAGAGCTGACCGCATTTCTTTGAAAAGATGATTGA 376
QY 582 AGAAGCTCGGTCATTCGGGATCAGATTTAATTTGATCCTTANGACTTCTGAAATGA 641
DB |||||
377 AGAGAAAGACAAAACCTCGGCAATGACCTGATTTGCTTTTGATCAAAAGCGCAACAAG 436
QY 642 AGGCATGGCGCTGTCGACAGAGATATCTCGCACTGATTTCTTAATGCTGTGTAGCGCG 701
DB |||||
437 AGGGGCAAGCTGTCCCTGTATGATGTTGCTTCCATTTTGAATCTGCTCTTGATGCGGG 496
QY 702 AACGGAACCGGCTGATGAAGACGCTGGCACTGATGATCTACCAATTTGCTCAACAATCCTGA 761
DB |||||
497 GAATGAACGACGACGAAATTAGTTTCAATGCGGCTACAGCATTTCTCGAAACACCGGG 556
QY 762 GCAGATGAATGATGTTTGGCTGACCGTTGCTGTTAGTTCCGAGAGCCATTTGCGGAGACAT 821
DB |||||
557 CGTGTATGACGAGCTCGCCAGGCATCTCGAATGATTTCCGCGGACGATCGAGGAGCGCT 616
QY 822 GCGTTATAACCGCGGTTGAGCTGATTCGCGGAGCTGTCCAGATACAGTGTGTCGG 881
DB |||||
617 CCGTTTCCGGGCGCCAGCCGATGATTTGCGGCTTCGTTAAACAGGATACCGAGATCAG 676
QY 882 CGGTATGGAATCAAAAAGATACGATGTTTGTGATGATCGGTGCGGCTTAAACCGGA 941
DB |||||
677 AGAGTAACTTGAATAAGAGAGAGTGTGATCGCTTTCTTGTCTTCTGCCAACCGTGA 736
QY 942 CCTGAAGCATTTGAACGCGCTGACGTGTTTAAATATTCATCG 983
DB |||||
737 TGAACGAAATTTAGAGGGCGCACGAATTTGATATTCACCG 778

RESULT 7

US-10-156-761-4523
; Sequence 4523, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4523
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1215)
US-10-156-761-4523

Query Match

5.6%; Score 67.6; DB 15; Length 1215;

Best Local Similarity 43.9%; Pred. No. 4.2e-10;
Matches 338; Conservative 0; Mismatches 429; Indels 3; Gaps 1;

QY 220 GGCCTGTGCTCGCCCAATGATGAAAAGAACACTCTGCCAAAAGAAAGAAATTTGATGTG 279
DB |||||
265 GACCATGGGATGCTGACCTGGAGCGCGGACCAACACCGGATCAGGCGCTGGTGTGCG 324
QY 280 AGAAGCTTTATCGGTGACGACATGATCATCTGTCTTCCATGATTAATAAATAAGTGAAG 339
DB |||||
325 AAGGGGTTTCAAGCGCGGACCGCTGGAGCAGCTCAAGCCGTAGCTGGCCAAAGCTCGCGCG 384
QY 340 AACTTTAGCGCTTATCTTGAAGAGGAAAGTGAATCTGTCATATTTTGAAGAG 399
DB |||||
385 GAGCTGTGGAACGGCTGCTCGCGGCGGCGGGGATCTGCTCGCGATGTGCGCGAG 444
QY 400 AGCTTTGCGGTGTGCTGACGATGACATGCTCGGGCTGGATAAAAAGAGACCATGAAAAA 459
DB |||||
445 CCGCTCCCGTGGCGTCACTCGCGAGATGCTGGGATCCCGAGTCCGACCGTGGCCCG 504
QY 460 ATCTCTGAGTGACACAGCGAGTGTGCGGATTTTATCAAGAGTATCTCTCAATCTCTGAA 519
DB |||||
505 CTCCGTCTGCTGCGCGGACATCTCGGGATGTACGAGTGAACCCGC---CGAAGGAC 561
QY 520 GCGCGGACATCTGTTATGTTGTCAGCGAAAGCTTTCCCAATCTTGTATCGCGTCAAT 579
DB |||||
562 GTGGCGGCAAGCGGTGCGGGCGTCTCGGAGTCTTCCGACTACCTGCGGAGCTGATC 621
QY 580 AAAGAACGTCGCGTCAATCCGGGATCAGATTTAAATTTTCGATCCTATGTACTTCTGAATAT 639
DB |||||
622 GCGAGGCGCGCAAGAGCGCGCGGACGATCTCCTCGGGCTCATCGCGCCCATGAC 681
QY 640 GAAGCATGGCGCTGTGCGACAAGGATATCTCGCATGATTTCTTAATGTGCTTTAGCC 699
DB |||||
682 GAAGCGACCGCTCACCGAGCAGGATGATCTCCACCTGCTGCTGCTCAACGCG 741
QY 700 GCAAACGAAACCGCTGATGAAGACGCTGACATGATGATCTACCATTTGCTCAACATCTT 759
DB |||||
742 GGCACGAGGCGCACCGTCAACGCCACGCTCAACGGCTGGTACGCTTGTTCGCAACCC 801
QY 760 GAGCAGATGAATGATGTTTGTGCTGACCGTCTGTTAGTTCCGAGAGCCATTTGGGAGACA 819
DB |||||
802 GACGAGTGGCGCGCTGGCGCGCCACCACTGCTGCTGCGCGCGCTGGAGGATGTG 861
QY 820 TTGCGTTATAACCGCGGTTTCAGCTGATTCGCGCGCAGCTGTCCCAAGATACAGTGGTC 879
DB |||||
862 ATGCGCTACGACACGCCCTCCAGCTCTTCGAGCGCTGGGTCTGGAACGAGATCGAGATC 921
QY 880 GCGGTATGGAATCAAAAAGATAGATGTTTGTGATGATCGGTGCGGCTAACCGG 939
DB |||||
922 GACGACGACGCTCCCGAGGGCGCGGAGATCGCCATGCTTTTGGCTCCGCCAACCCAC 981
QY 940 GACCTGGAAGCATTTGAACAGCGCTGACGCTGTTTAATATTCATCGGGAAGA 989
DB |||||
982 GACCCCGAGTCTTCGGGATCCCGAGAAATCGACCTCACCCCGGAGGA 1031

RESULT 8

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 5.6%; Score 67.6; DB 15; Length 9025608;
Best Local Similarity 43.9%; Pred. No. 1.2e-07; Mismatches 429; Indels 3; Gaps 1;
Matches 338; Conservative 0; Mismatches 429; Indels 3; Gaps 1;
QY 220 GGCCCTGTGCTGCGCCCAATGATGAAAGAACACTCTGCCAAAGAGAAATTTGTAGTG 279
DB 5543641 GACCATGGATGCTCGACCTGGAGCCGCGGACACACCCGGATCAGGCGCTGTGTGCG 5543700
QY 280 AGAAGCTTTATCGGTGACGACTGGATCATCTGTCTCCATTGATTAACAAATGAGAA 339
DB 5543701 AAGGCGTTTACGCGCGGACCGTGGAGCAGCTCAAGCCGTAGTGGCCAAAGCTCGCCGC 5543760
QY 340 AACTTTGAGCGCTTATCTTGAAGAGGAAAGTGTCTCGTCAATGATTTTGAAG 399
DB 5543761 GAGCTGGTGGACCGGCTGCTCGCGCGGCGCGGGGATCTGCTCGCGGATGTCCCGAG 5543820
QY 400 AGCTTTGCGGTGTGCTGACGATGACATGCTCGGGCTGGATAAAGAGACCATGAAAA 459
DB 5543821 CCCTCCCGTGGCGGTGATCGCGGAGATGCTGGGATCCCGAGTCCGACCGTGCCCG 5543880
QY 460 ATCTCTGAGTGACACAGCGAGGTGCGGATTTTATCAAGAGTATCTCTCAATCTCTGAA 519
DB 5543881 CTCCGTCCCTGCTCGCGCGGACATCTCGGGATGTACGAGCTGAAACCGG---CGAAGGAC 5543937
QY 520 GCGCGGACATTCGTTATGTTGCGAGCAACAGCTTTCCCAATCTTGTATGCGGTCATT 579
DB 5543938 GTGGCGGCGAAGCGGCTGCGGGCGTGGTGGAGTCTTCGAGTACTCTGCGGAGTGTATC 5543997
QY 580 AAGAAGCTGCGGTCAATCGGGATCAGATTTAAATTTTCGATCTATGATCTCTGAATAT 639
DB 5543998 GCGGAGCGCGCAAGAGCGCGGCGAGATCTCATCTCGGGCTCATCGCCGCCATGAC 5544057
QY 640 GAAGCATGCGCTGTGGAACAAGGATATCTCGCACTGATTTCTTAATGTGCTGTAGCC 699
DB 5544058 GAAGCGACCGCTCACCGGAGGAGATGATCTCCACCCTGCTGCTGCTCAACGCG 5544117
QY 700 GCAACGGAACCGCTGATGAAGCGTGGCACTGATGATACCATTTGCTCAACATCTT 759
DB 5544118 GGCCACGAGGCCACCGTCAACGCCACCGGTCAAACGGTGTAGCCCTTGTTCGCAACCCC 5544177
QY 760 GAGCAGATGAATGATGTTTTTGGTGTACCGTTCGTTAGTTCCGAGAGCCATTCGGGAGCA 819
DB 5544178 GACCACTGGCGCCCTGCGCGCGGACCACTGCTGCTGCGCGCCGCTGGAGGTTG 5544237
QY 820 TTGCGTTATAACCGCGGTTGAGTGTATTCGCGGCGAGCTGTCCCAAGATACAGTGTGTC 879
DB 5544238 ATGCGCTACGACACGCGCTCCAGCTCTTCGAGCGGTGGTCTGAGCAGATCGAGATC 5544297
QY 880 GCGGTATGGAATCAAAAAGATAGATGTTTTTTGATGATGATCGGTGCGGCTAACCGG 939
DB 5544298 GACGGCACACCGTCCCGAGGGCGCGGAGATGCGCATGCTTTTGGCTCCGCCAACCCAC 5544357
QY 940 GACCCCTGAAGCATTTGAACAGCCTGACGTTGTTAATATTATTCGCGGAAGA 989
DB 5544358 GACCCCGAGGTCTTCCGGATCCGAGAACTCGACCTCACCCGCGAGGA 5544407

RESULT 9
US-09-974-300-2182

; Sequence 2182, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2182
; LENGTH: 1213
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2182

Query Match 5.3%; Score 64.8; DB 9; Length 1213;
Best Local Similarity 47.0%; Pred. No. 3.3e-09;
Matches 201; Conservative 0; Mismatches 227; Indels 0; Gaps 0;
QY 560 AATCTTCATGCGGTCTAATAAGAACGTCGCGTCAATCCGGGATCAGATTTAATTTGCA 619
DB 589 ATTATTTTAAACATTTATCAGCAGCGCAAAAAGAGCCAAAAGACGACCTGATTTTCG 648
QY 620 TCCTATGTACTTCTGAATATGAAGCATGGCGCTCTCGGACAAAGGATATCTCGCACTGA 679
DB 649 TTTTACTCGGGCGGAAGTTGACGCGCAATCGTCACAGAGAAGAACTGCTTCATTTT 708
QY 680 TCTTAATGTGCTGTAGTCGCAACGGAACCGGCTGATAAGACGCTGGCACTGATGATCT 739
DB 709 GCATCATTTTGTGTCGAGCAATGAGACGACAACTTGATCGCAAAACAGCGTCC 768
QY 740 ACCATTTCCTCAACAACTCTGAGCAGATGAATGATTTTGGCTGACCGTTCGTTAGTTC 799
DB 769 GCTATCTCAGAGAAGATAAATTCACAGGAAGCCGTGAAGACAAGATCGTCCCTGCTCC 828
QY 800 CGAGAGCCATTCGCGGAGACATTCGCTTATAAACCCCGGTTCCAGCTGATTCGCGGCGAC 859
DB 829 CTGCTCTTTTGAAGAAATGCTGCGTTATTATCCCGCGTGCAGCGATCGCGCGCACGG 888
QY 860 TGTCACAGATACAGTGTGCGCGGTATGGAATCAAAAAGATAGGATGTTTTTTGTA 919
DB 889 CGGCAGAAAGACGTTGATATCGGAGCGTGAAGATTGCAAAAGGTTCTACAGTGAATCAGCT 948
QY 920 TGATCGGTGCGGCTAACCGGACCCCTGAAGCATTTGAAACAGCTGACGTTTAAATATTC 979
DB 949 GGGTCTGCTTACGGAATCTGTACGAACCTTAAGTTTGAAGATCTCTGACAGCTTCAAGCTTG 1008
QY 980 ATCGGGA 987
DB 1009 ATCGCAA 1016

RESULT 10
US-10-282-122A-25888
; Sequence 25888, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 25888
LENGTH: 1200
TYPE: DNA
ORGANISM: Mycobacterium avium
US-10-282-122A-25888

Query Match 4.6%; Score 55.6; DB 16; Length 1200;
Best Local Similarity 45.6%; Pred. No. 2.8e-06;
Matches 196; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY	560	ATACTTGATCGCGTCAATTAAGACGTCGCGTCAATCGGGATCAGATTAAATTCGA	619
Db	575	ACTACTTCTCGAGCAGATCGAACAACGCAAGTCCAAAGCCACCGAGGACATCATCGGGG	634
QY	620	TCCTATGTACTTCTGAATATGAAGCATCGCGTGTGCGACAAGGATATACCTCGCACTGA	679
Db	635	ACCTGTCACCGGGAGATCGACGGCGAAAGCTCAGCAGCAGGCGCATCTACTCGTTCC	694
QY	680	TTCTTAATGTGTGTAGCCGCAACGGAAACCGGCTGATGAAGACGCTGGCACTGATGATCT	739
Db	695	TGCGGCTGCTGTTGCCGCGCGGGCTGAAACACCACTACCGGTCTGCGGAAACCTGTTGT	754
QY	740	ACCATTTGCTCAACATCTTGACGACAGATGAATGATGTTTGGCTGACCGTTGTTAGTTC	799
Db	755	ATCTGCTGCTACCCACCGCGGACCAAGTTCCGCGCTGCGAGCGGATCGGAGCTGCTGG	814
QY	800	CGAGAGCCATTGCGGAGACATTGCGTTATAAACCGCGGTTCAGCTGATTCCGCGGCAGC	859
Db	815	CGCCGCGGATCAGGAGGGGTTGCGGTTTCGAGACACCGCTGACCCGCTGCGAGCGTTCA	874
QY	860	TGTCCCAAGATACAGTGGTTCGCGGATGGAATAAAAAAGATAGATTGTTTTTGTGA	919
Db	875	CCACCGAAGACACCGAACTGCACGGGTCGGATTCCGGCGGCTTCGGTGATCGCGGTGT	934
QY	920	TGATCGGTCGGGTAAACCGGACCCCTGAAGCATTTGAACAGCCTGACGTTGTTAATTC	979
Db	935	GCATCGGTTCCGGAACCCCGACGAGCGGCTGGGAGCGTTCCGAGGAATTCGACATCT	994
QY	980	ATCGGGAACA 989	
Db	995	TCCGTAACA 1004	

RESULT 11
US-10-229-148B-1/c
Sequence 1, Application US/10229148B
Publication No. US20040091975A1
GENERAL INFORMATION:
APPLICANT: Meiji Seika Kaisha, Ltd.
TITLE OF INVENTION: Midecamycin biosynthetic genes
FILE REFERENCE: 138451 US
CURRENT APPLICATION NUMBER: US/10/229,148B
CURRENT FILING DATE: 2002-08-28
PRIOR APPLICATION NUMBER: 210516/2002
PRIOR FILING DATE: 2002-07-19
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 84428
TYPE: DNA
ORGANISM: Streptomyces mycarofaciens
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((1)..(675))
OTHER INFORMATION: ORF42 (fragment)
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((1168)..(2202))
OTHER INFORMATION: ORF41
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((2220)..(3215))
OTHER INFORMATION: ORF40
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((3237)..(4691))
OTHER INFORMATION: ORF39
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((4695)..(5948))
OTHER INFORMATION: ORF38
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((6048)..(6629))
OTHER INFORMATION: ORF37
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((6653)..(7945))
OTHER INFORMATION: ORF36
FEATURE:
NAME/KEY: CDS
LOCATION: (8149)..(9015)
OTHER INFORMATION: ORF35
FEATURE:
NAME/KEY: CDS
LOCATION: (9012)..(9335)
OTHER INFORMATION: ORF34
FEATURE:
NAME/KEY: CDS
LOCATION: (9328)..(10458)
OTHER INFORMATION: ORF33
FEATURE:
NAME/KEY: CDS
LOCATION: (10521)..(11603)
OTHER INFORMATION: ORF32
FEATURE:
NAME/KEY: CDS
LOCATION: (11729)..(12961)
OTHER INFORMATION: ORF31
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((13016)..(14044))
OTHER INFORMATION: ORF30
FEATURE:
NAME/KEY: CDS
LOCATION: Complement ((14074)..(15096))

```

; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((75899)..(76570))
; OTHER INFORMATION: ORF10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((76602)..(77765))
; OTHER INFORMATION: ORF11
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78039)..(79313)
; OTHER INFORMATION: ORF12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((79391)..(81052))
; OTHER INFORMATION: ORF13
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81541)..(82356)
; OTHER INFORMATION: ORF14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (82760)..(83362)
; OTHER INFORMATION: ORF15
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((83495)..(84142))
; OTHER INFORMATION: ORF16
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (84329)..(84428)
; OTHER INFORMATION: ORF17 (fragment)
; US-10-229-148B-1

Query Match      4.6%; Score 55.6; DB 16; Length 84428;
Best Local Similarity 45.4%; Pred. No. 4.1e-05;
Matches 199; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

QY 546 CGAACAGCTTTCCCAATCTTGATCCGGTCAATTAAAGAACGTGCGTCAATCCGGGATC 605
DB 26822 CAAGGAGATGACCGACTATCTCTGCACCGATCCACGCAAGCGAGGCCCGGTGGA 26763
QY 606 AGATTTAATTTCGATCCTTACTTCTGAATATGAAGCATGGCGCTGCGGACAAGGA 665
DB 26762 CGACCTGATCAGCGGCTGGTCCAGGCGGAGCAGACGCCGCAAGCTCACCAGCTGGA 26703
QY 666 TATACTCGCACTGATTCTTAATGTGCTGTAGCCCAACGGAACGGGCTGATAAGACGCT 725
DB 26702 GATCGTCAACATCGTGGCCCTCTGCTCACCGCCGCCACGCTCTGTCAGCACGCTGT 26643
QY 726 GGCACGTGATGATCTACCAATTTGCTCAACAATCCTGAGCAGATGAATGATGTTTGGCTGA 785
DB 26642 CAGCAACCTCTTCTCTGGTACTGAGGAGAAACCGCAGGCGCTGGCGGACCTGCGCCGA 26583
QY 786 CCGTTTCGTTAGTTCGAGAGCCATTGCGGAGACATTGCGTTATAACCCCGGTTTCAGCT 845
DB 26582 CCGGAGCTGGTGAACCGGTGGCGTGGAGGAGACACTGGCTACCGCAGCCCTTCAACAA 26523
QY 846 GATTCCGGCGGACGCTGCCAAGATACAGTGTGTCGCGCTATGGAATCAAAAAAGATAC 905
DB 26522 CATCTTCGCTTCCTCAAGGAGGACACCGACATCTCGGCCCGGAATGAAGAGGCCCA 26463
QY 906 GATTGTTTTTTGATGATCGGTGCGGCTAACCGGAGACCTCTGAAGCATTTTGAACAGCTGA 965
DB 26462 GATGTCATCGCTCGGAGCCAGTCGCGCAACCGCGCAACCTTCCCGAGGCCGA 26403
```

QY 966 CGTGTATTAATTCATCG 983
Db 26402 CACCTTCGACATCCGAGC 26385

RESULT 12

US-10-205-032-7
; Sequence 7, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-7

Query Match 4.5%; Score 55.2; DB 15; Length 1233;
Best Local Similarity 45.3%; Pred. No. 3.8e-06;
Matches 201; Conservative 0; Mismatches 243; Indels 0; Gaps 0;
QY 543 CAGCGAACAGCTTTCCCAATACCTTGATGCGCGTCATTAAGAACGTCGCGTCAATCCGGG 602
Db 576 CATCGAGGAGTGTCCGAATTCCTGATCCGAGATCGCCACAAAGCGCGCGAACCCGAA 635
QY 603 ATCAGATTAAATTCGATCCTATGTAATCTTGAATATGAGGATCGGCGCTGTCGACAA 662
Db 636 GGACGACCTGATCAGCGGCTCTGTGCGGCGAGGTGCGAGCGGCGCAAGCTGACCGACGA 695
QY 663 GGATATACCTCGACATGATTTAAATGCTGTGTAGCGCGAACGGAACCGGCTGATAAGAC 722
Db 696 GGAAGTCGTCAACATCGTCGCGTGTCTACCGCGCGGCACATCTCAGCGCCACGCT 755
QY 723 GCTGCGACCTGATGATACATTTGCTCAACATCTCGAGCAGATGAATGATGTTTGGC 782
Db 756 GCTCAGCAACCTGTTCTGCTGTGAGGAGCACCGCGAGGACAGCGCGCGTCCGCGC 815
QY 783 TGACCGTTCGTTAGTTCGAGAGCCATTGCGGAGACATTCGCTTATAAACCGCGGTTCA 842
Db 816 CGACCGCAGCCTCGTCCGCGGCGTGTGATCGAGGAGACGCTGCGCTACCGGTCGCCGTTCAA 875
QY 843 GCTGATTCGCGCGACGCTGCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGA 902
Db 876 CTGATCTTCGGATCCTGAAACGAGACACCGACATCTCTGCGCCACCCCATCGGCAAGG 935
QY 903 TACGATGTTTTTGTATGATCGGTGCGGTCAACCGGACCCCTGGAAGCATTTTGAACAGCC 962
Db 936 CCAGATGTTGATCGCTGATCGCTCCCGAACCGGACACCGAGGTGTTTACCGGACCC 995
QY 963 TGACGTGTTTAATTCATCGGA 986
Db 996 GGACACCTTCGACATCCGAGCGGA 1019

RESULT 13

US-10-205-032-1/c
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US

; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-1

Query Match 4.5%; Score 55.2; DB 15; Length 60196;
Best Local Similarity 45.3%; Pred. No. 4.5e-05;
Matches 201; Conservative 0; Mismatches 243; Indels 0; Gaps 0;
QY 543 CAGCGAACAGCTTTCCCAATACCTTGATGCGCGTCATTAAGAACGTCGCGTCAATCCGGG 602
Db 3286 CATCGAGGAGTGTCCGAATTCCTGATCCGAGATCGCCACAAAGCGCGCGAACCCGAA 3227
QY 603 ATCAGATTAAATTCGATCCTATGTAATCTTGAATATGAGGATCGGCGCTGTCGACAA 662
Db 3226 GGACGACCTGATCAGCGGCTCTGTGCGGCGAGGTGCGAGCGGCGCAAGCTGACCGACGA 3167
QY 663 GGATATACCTCGACATGATTTAAATGCTGTGTAGCGCGAACGGAACCGGCTGATAAGAC 722
Db 3166 GGAAGTCGTCAACATCGTCGCGTGTCTACCGCGCGGCACATCTCAGCGCCACGCT 3107
QY 723 GCTGCGACCTGATGATCTACCAATTTGCTCAACATCTCGAGCAGATGAATGATGTTTGGC 782
Db 3106 GCTCAGCAACCTGTTCTCTGCTGAGGAGCACCGCGAGGACAGCGCGGTCGCGC 3047
QY 783 TGACCGTTCGTTAGTTCGAGAGCCATTGCGGAGACATTCGTTATAAACCGCGGTTCA 842
Db 3046 CGACCGCAGCCTCGTGTGCGGCGTGTATCGAGGAGACGCTGCGCTACCGGTCGCCGTTCAA 2987
QY 843 GCTGATTCGCGCGACGCTGCCAAGATACAGTGTGCGCGGTATGGAATCAAAAAGA 902
Db 2986 CTGATCTTCGGATCTTGAACGAGACACCGACATCTCTGCGCCACCCCATCGGCAAGG 2927
QY 903 TACGATGTTTTTGTATGATCGGTGCGGCTAACCGGACCCCTGGAAGCATTTTGAACAGCC 962
Db 2926 CCAGATGTTGATCGCTGATCGCTCCCGAACCGGACACCGAGGTGTTTACCGGACCC 2867
QY 963 TGACGTGTTTAATTCATCGGA 986
Db 2866 GGACACCTTCGACATCCGAGCGGA 2843

RESULT 14

US-09-974-300-2185
; Sequence 2185, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2185
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2185

Query Match 4.5%; Score 55; DB 9; Length 1155;
Best Local Similarity 48.0%; Pred. No. 4.2e-06;

Matches 157; Conservative 0; Mismatches 170; Indels 0; Gaps 0;			
QY	651	GCTGTCGACAGGATATACCTCGACTGATTTCTTAATGTCGTGTAGCCGCAACGGAACC	710
Db	637	GCTTGCTGATGAAGAAGTCTTGCGACGFGCATCTCGCTTGTGATTCGCCGACATGAGAC	696
QY	711	GGCTGATAAGACGCTGGCACTGATGATCTACCATTTGCTCAACAATTCCTGAGCAGATGAA	770
Db	697	GACAGTCAACCTTTTGACCAACGGGTGTTTGTGCTCATGACATCTCTGACCAGCTCAG	756
QY	771	TGATGTTTGGCTGACCGTTCGTTAGTTCGAGAGCCATTCGGGAGACATTCGCTTATAA	830
Db	757	TGAGCTGAAAGAAAATCCGCTGCTGATTCAGTCGCCGTCGAAGAATGCTTCGCTATGA	816
QY	831	ACGCGCGTTCAGCTGATTCGCGGCGAGCTGTCCCAAGATACAGTGGTCGGCGGTATGGA	890
Db	817	AAGCCCGACACAGCTGACAGCGCGCACCGCTCGGAGGACTCGCAAAATCAACGGGAAAAT	876
QY	891	AATCAAAAAAGATACGATTTGTTTGTATGATCGGTGCGGCTAACCGGGAGCCCTGAAGC	950
Db	877	CATCAAAAAAGGAGAGCACTTATATATCTTTTAGCGCGGCCACCGCGATCCAAAAGAT	936
QY	951	ATTGAACAGCCTGACGTGTTTAATAT	977
Db	937	ATTTCAAAACCGCATCTATTGACAT	963

RESULT 15
US-10-214-446-49
; Sequence 49, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 1215
; TYPE: DNA
; ORGANISM: Bacterial
US-10-214-446-49

Query Match 4.4%; Score 53.6; DB 15; Length 1215; Best Local Similarity 45.6%; Pred. No. 1.2e-05; Matches 168; Conservative 0; Mismatches 224; Indels 0; Gaps 0;			
QY	560	ATATCTTGATGCGGTTCATTAAGAACGTCGCGTCAATCCGGGATCAGATTTTAATTCGA	619
Db	581	ACTACTTCGCCGACTCTCTGCGGAGCGGGAGAAACCCGCGGACGACCTGGTCAGCC	640
QY	620	TCCTATGTAATCTGAATATGAGGCATGCGCTGTCGAGCAAGGATATACTCGCACTGA	679
Db	641	TGCTGCTAGCCCGGAGCGGAGGGGGAGTGATGACCCGCGAGGTCCTGGGGAAT	700
QY	680	TTCTTAATGTGCTGTTAGCCGCAACCGGCTGATAGACGCTGGCACTGATGATCT	739
Db	701	GCACCCAGATCATCGTCGCGGACAGACACCCGCACTCGTGGCCACGGGCTCG	760
QY	740	ACCATTTGCTCAACAAATCTCGAGCAGATGAATGATGTTTGGCTGACCGTTGTTAGTTC	799
Db	761	AACTCCTCCTCGGTACCCGCGAGCAGCGGCCCTCTCGGAATCCCGCCGAGAGTGATGC	820

Search completed: December 10, 2004, 21:11:10
Job time : 768.715 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:45:13 ; Search time 135.823 Seconds
(without alignments)
1069.666 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQIKLPSVLSDQFNPNP.....AESGLYTRGPVSLVAFDGA 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 359729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	405	8	ADJ62039 Bacillus
2	958	46.1	408	6	ABM70363 Photorhab
3	454	21.8	410	6	AAE36997 Micromono
4	440.5	21.2	399	6	ABP57702 Saccharop
5	438	21.1	403	8	ADM80057 Spiramyci
6	438	21.1	403	8	ADN97573 S ambofac
7	433.5	20.8	404	8	ADH39706 Streptomy
8	432	20.8	403	8	ADM45932 S. mycaro
9	430.5	20.7	404	6	ABR82126 Bacterial
10	426.5	20.5	396	3	AAE15502 Bacillus
11	423.5	20.4	404	6	ABR82143 Bacterial
12	418.5	20.1	396	3	AAE15501 Bacillus
13	417.5	20.1	396	3	AAE15503 Bacillus
14	415.5	20.0	396	2	AAE60777 Mycinamic
15	413	19.9	395	6	ABR82124 Bacterial
16	412.5	19.8	408	6	ABR82120 Bacterial
17	406	19.5	392	6	ABR82134 Bacterial
18	399.5	19.2	425	6	ABR82128 Bacterial
19	394.5	19.0	406	8	ADH39695 Streptomy
20	393	18.9	397	7	ADJ53176 Rebecamy
21	392.5	18.9	400	6	ABU34148 Protein e
22	391	18.8	418	6	AAE35502 Streptomy
23	383.5	18.4	402	6	ABU11981 M. echino
24	374	18.0	430	6	AAE34780 Streptomy
25	373.5	18.0	396	5	ABR81315 Streptomy

26	373.5	18.0	430	6	AAE34736 Streptomy
27	371.5	17.9	430	6	AAE34735 Streptomy
28	371.5	17.9	430	6	AAE34723 Streptomy
29	366.5	17.6	430	6	AAE34727 Streptomy
30	366.5	17.6	430	6	AAE34725 Streptomy
31	366.5	17.6	430	6	AAE34731 Streptomy
32	365.5	17.6	404	6	ABR82138 Bacterial
33	365	17.5	429	6	AAE34728 Streptomy
34	364.5	17.5	403	7	ADD15221 Eryf prot
35	364.5	17.5	403	8	ADO47230 S. erythra
36	362.5	17.4	424	8	ADM45925 S. mycaro
37	361.5	17.4	424	8	ABR82146 Bacterial
38	361	17.4	475	6	AAE34729 Streptomy
39	360	17.3	416	3	AAE18654 Amino aci
40	360	17.3	416	3	AAE18654 Macrolide
41	360	17.3	416	6	ABG71678 S. venezu
42	360	17.3	416	6	ADA09417 S. venezu
43	360	17.3	416	7	ADH53461 S. venezu
44	360	17.3	416	8	ADL91924 Streptomy
45	356.5	17.1	418	8	ADL39276 S. hygro

ALIGNMENTS

RESULT 1
ADJ62039
ID ADJ62039 standard; protein; 405 AA.

XX AC ADJ62039;
XX DT 06-MAY-2004 (first entry)
XX DE Bacillus subtilis cypX protein SEQ ID NO:2.
XX KW heterologous biological substance; Bacillus; cypX; red pigment.
XX OS Bacillus subtilis.

XX PN WO2004011609-A2.
XX PD 05-FEB-2004.
XX PF 25-JUL-2003; 2003WO-US023398.
XX PR 26-JUL-2002; 2002US-0398853P.

XX PA (NOVO) NOVOZYMES BIOTECH INC.
XX PI Tang M, Sloma A, Sternberg D, Behr R;
XX DR WPI: 2004-143839/14.
XX DR N-PSDB; ADJ62038.

XX PT Producing a heterologous biological substance comprises cultivating the pigment-deficient mutants of Bacillus cell that directs synthesis of the heterologous biological substance and has a modification of the cypX and ymcC genes.

XX PS Example 1; SEQ ID NO 2; 62pp; English.
XX CC The present invention describes a method for producing a heterologous biological substance comprising cultivating a mutant of a parent Bacillus cell in a medium suitable for the production of a heterologous biological substance, and recovering the heterologous biological substance from the cultivation medium, where the mutant cell comprises a first nucleic acid sequence directing synthesis of the heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes cypX and ymcC, which are involved in the production of a red pigment. The mutant cell is also deficient in the production of the red pigment compared to the parent Bacillus cell when cultivated under the same conditions. Also described: (i) a mutant of a parent Bacillus cell, comprising a first nucleic acid sequence directing synthesis of a

Adi-1003

heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes *cypX* and *yvmC*, which are involved in the production of a red pigment, wherein the mutant cell is deficient in the production of the red pigment compared to the parent *Bacillus* cell when cultivated under the same conditions; and (2) obtaining a mutant of a parent *Bacillus* cell, comprising introducing into the parent *Bacillus* cell a first nucleic acid sequence directing synthesis of a heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes *cypX* and *yvmC*, which are involved in the production of a red pigment, and identifying the mutant cell comprising the modified nucleic acid sequence, where the mutant cell is deficient in the production of the red pigment compared to the parent *Bacillus* cell when cultivated under the same conditions. The methods and compositions of the present invention are useful for producing biological substances, e.g. nucleic acid, polyamide, polyamine, polyol, polypeptide, or polysaccharide, in pigment-deficient mutants of *Bacillus* cells, and obtaining the pigment-deficient mutant cells. The present sequence represents *Bacillus subtilis cypX*, which is used in the exemplification of the present invention.

XX Sequence 405 AA;

Query Match 100.0%; Score 2080; DB 8; Length 405;
Best Local Similarity 100.0%; Pred. No. 1e-196;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSQSILFSLSDQFQNNPYAFSOLREDPVHYEESIDSYFISYHVDVRYILOHPDIPT 60
DB 1 MSQSILFSLSDQFQNNPYAFSOLREDPVHYEESIDSYFISYHVDVRYILOHPDIPT 60
QY 61 TKSILVERAEPVNRGPPVLAQMHGKHSKRRIIVRSGFIGDALDHLSPLIKQNAENLLAPYL 120
DB 61 TKSILVERAEPVNRGPPVLAQMHGKHSKRRIIVRSGFIGDALDHLSPLIKQNAENLLAPYL 120
QY 121 ERGKSDLVNDFOKTFVAVCVTMQMLGDKRDHKIKSEWHSGVADFTISISQSPARAHSIW 180
DB 121 ERGKSDLVNDFOKTFVAVCVTMQMLGDKRDHKIKSEWHSGVADFTISISQSPARAHSIW 180
QY 181 CSEQLSOYLMPIVKERRVNPQSDLSILCTSEYEGMALSDKDILALINVLAAATEPADK 240
DB 181 CSEQLSOYLMPIVKERRVNPQSDLSILCTSEYEGMALSDKDILALINVLAAATEPADK 240
QY 241 TLALMIYLLNPNPEQNDVLADRSILVPRAIATLRYKPPVQLIPQLSQDVTVGMEIKK 300
DB 241 TLALMIYLLNPNPEQNDVLADRSILVPRAIATLRYKPPVQLIPQLSQDVTVGMEIKK 300
QY 301 DTIVFCMIGAAVRDPEAFQPDVFNHREDLGKSAFSGAARHLAFGSGIHNCVGAFAK 360
DB 301 DTIVFCMIGAAVRDPEAFQPDVFNHREDLGKSAFSGAARHLAFGSGIHNCVGAFAK 360
QY 361 NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405
DB 361 NEIEIVANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405

RESULT 2

ABW70363

ID ABW70363 standard; protein; 408 AA.

XX

AC ABW70363;

XX 20-NOV-2003 (first entry)

XX Photorhabdus luminescens protein sequence #3460.

DE Photorhabdus luminescens protein sequence #3460.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

KW detection; food; gene expression; plant; animal; microorganism; toxin;

KW antibiotic; biopesticide; virulence factor; disease model; plague;

KW whooping cough.

XX Photorhabdus luminescens.

OS Photorhabdus luminescens.

XX WO200294867-A2.

PN

XX 28-NOV-2002.
PD 07-FEB-2002; 2002WO-IB003040.
XX 07-FEB-2001; 2001FR-00001659.
PR (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
PA Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX WPI; 2003-148459/14.
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
PT Claim 2; SEQ ID NO 3460; 1205pp; French.
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX

Sequence 408 AA;

Query Match 46.1%; Score 958; DB 6; Length 408;
Best Local Similarity 46.6%; Pred. No. 1.3e-85;
Matches 185; Conservative 82; Mismatches 124; Indels 6; Gaps 3;

QY 8 FSVLSDDQFQNNPYAFSOLREDPVHYEESIDSYFISYHVDVRYILOHPDIPTTKSLVER 67
DB 7 FNHDPKFKIPYDFYDILHKQDLVYFQSQNSYFQYGVKEDVDALK-SSIENTKPLTAL 65
QY 68 AEPVNRGPPVLAQMHGKHSKRRIIVRSGFIGDALD-HLSPLIKQNAENLLAPYLRGKSD 126
DB 66 AEPVNRGPPVLAQMHGKHSKRRIIVRSGFIGDALD-HLSPLIKQNAENLLAPYLRGKSD 125
QY 127 LVNDPGKTFVAVCVTMQMLGDKRDHKIKSEWHSGVADFTISISQSPARAHSIWSEQLS 186
DB 126 LVNDPGKTFVAVCVTMQMLGDKRDHKIKSEWHSGVADFTISISQSPARAHSIWSEQLS 185
QY 187 QYLMPIVKERRVNPQSDLSILCTSEYEGMALSDKDILALINVLAAATEPADKTLALMI 246
DB 186 RLLKPIIDORRNPQSDLSILCTSEYEGMALSDKDILALINVLAAATEPADKTLALMI 241
QY 247 YHLLNPNPEQNDVLADRSILVPRAIATLRYKPPVQLIPQLSQDVTVGMEIKKDTIVFC 306
DB 242 NHHLSNPSMLDVLKDRSLVRAFAEETLRLTSPVQLIPREASEDVTISGIDIPKGAWVFC 301
QY 307 MIGAARDPEAFQPDVFNHREDLGKSAFSGAARHLAFGSGIHNCVGAFAKKEIIV 366
DB 302 MIGAARDPEAFQPDVFNHREDLGKSAFSGAARHLAFGSGIHNCVGAFAKKEIIV 361
QY 367 ANIVLDKMRNIRLEEDFCYAESGLYTRGPVSLVAFD 403


```
Db 64 AAAILNRDVRGFPILRLREHSLGTMDDPPEHTRRLKLVGKAFARRVEQLRPRQLQVMDHL 123
QY 116 LAPYLERG-KSDIYNDPGKTFVAVCVTMDMLGDKRDHEKISWHSGVADFTISQSPEA 174
Db 124 LORMAADGPGDLVSALALPLPIKVICDLGIPVADRFRVW-SDIALAIT--SNSPEE 180
QY 175 RAHSLWCSQLSQYLMPVIKERRVNPGLISILCTSEYEGMALSDKOILALILNVLAA 234
Db 181 IRES---RQIIRAYIGELVQQRKKPTEDLLSVLQVRAEGQLSEELVITGAGLLIAG 237
QY 235 TEPADKTLALMIYHLNPNQMDVADRSLVPRAIAETLYKP--PVQLIPRQLSQDQTV 292
Db 238 FETTANHIANFTFNLLTHPDQLKLIADPELVRAVEELLRYTLGATPGFPRIATEDLE 297
QY 293 VCGMEIKKDTIVFCMIGANRDPFAEQPDVFNHREDLGKISAFSGAARHLAFSGIHN 352
Db 298 LGGVSIIRGDVAFVFEIASANRDSAVFDGPDLDLAREH-----NSHVALGHGPHY 347
QY 353 CVGAFAKNEIEIVANIVLDKMRNI--RLEEDFCVAESGLYTRGVPVSLLVAF 402
Db 348 CIGQLARWELQVAILTKRFPQLSFVAVPDEVVWVWKRGMTRGPEALPITW 399

RESULT 5
ADM80057
ID ADM80057 standard; protein; 403 AA.
AC ADM80057;
XX
XX
DT 01-JUL-2004 (first entry)
DE Spiromycin biosynthesis orf1 protein, SEQ ID 24.
KW Antibacterial; Immunosuppressive; Antirheumatic; Antiarthritic;
KW Anthelmintic; Insecticide; spiramycin biosynthesis; macrolide;
KW spiramycin; antibiotic; immunosuppressant; autoimmune disease.
XX
XX OS Streptomycetes ambofaciens.
XX
XX PN FR2845394-A1.
XX
XX PD 09-APR-2004.
XX
XX PF 08-OCT-2002; 2002FR-00012489.
XX
XX PR 08-OCT-2002; 2002FR-00012489.
XX
XX PA (AVET ) AVENTIS PHARMA SA.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX PI Blondelet RMH, Dominguez H, Darbon RE, Gerbaud C, Gondran A;
XX PI Karray F, Lacroix P, Oestreicher MBN, Pernodet JL, Tuphile K;
XX
XX DR WPI: 2004-307291/29.
XX N-PSDB; ADM80056.
XX
XX PT New polynucleotides encoding proteins involved in spiramycin
XX biosynthesis, useful for improving synthesis of macrolide antibiotics or
XX for generating new hybrid macrolides.
XX
XX PS Claim 8; SEQ ID NO 24; 323pp; French.
XX
XX CC The present invention relates to coding sequences (I) which encodes
XX proteins involved in spiramycin (A) biosynthesis. (I) are used to improve
XX CC production (optionally also purity) of macrolides (M), especially
XX CC spiramycins, and to produce hybrid antibiotics. (M) are useful in human
XX CC and veterinary medicine as antibiotics, but some are also
XX CC immunosuppressants (e.g. FK506) useful for treating transplant rejection,
XX CC rheumatoid arthritis and other autoimmune diseases; or insecticides and
XX CC anthelmintics (e.g. avermectin). The present sequence is one such
XX protein involved in spiramycin biosynthesis.
XX
```

```
SQ Sequence 403 AA;
Query Match
Best Local Similarity 21.1%; Score 438; DB 8; Length 403;
Matches 110; Conservative 77; Mismatches 156; Indels 34; Gaps 7;
QY 22 YFSQLREDDPVHYEESIDSYFISRYHVRVYILOHPDIFTTK-SLVERAEPVNR---GP-V 76
Db 23 WFAYNRTRHPVFWDETRQAWQVGYDDVYTVSNPNPQFFSSDFNMVMPPTPELEMIIGPOT 82
QY 77 LAQMHGKHSKRRIRVRSFISGDLALDHLSPILKQNAENLLAPYLERGKSDLVNDFGKTPA 136
Db 83 IGALDPPAHGPMKLVLSQAFTPRRIARLEPRVRAITEELLKVGQDDVVDVADGDLALP 142
QY 137 VCVTMDMLGDKRDHEKISEW-----HSGVADFTISQSPEARAHSLWCSEOL 185
Db 143 VIVIAELGIPAGRDLDLFWWDVTLTNEGLEYPLNDPNT-----ETIAPAL---KEM 193
QY 186 SQYLMFVIKERRVNPGLISILCTSEYEGMALSDKOILALILNVLAAATEPADKTLALM 245
Db 194 TDYLLKQIHAKRDAPADLVSLVQAEQDGRRLTDEIVNIVALLLTAGHVSSTLLSNL 253
QY 246 IYHLNPNQMDVADRSLVPRAIAETLYKPVLIPRQLSQDQTVVCGMEIKKDTIVF 305
Db 254 FLVLEENPQALDLRADRSLVFGAIEETLYRSPFNNTFRFYKEDTTVLGPLMERGQWVI 313
QY 306 CMIGAANRDPFAEQPDVFNHREDLGKISAFSGAARHLAFSGIHNCGVAFAKNEIEI 365
Db 314 AWSQSANRDPHFDPDPTDFIRSD-----GTRHWAFGHGHHCGLGAALARLEGV 364
QY 366 VANIVLDKVRNIRLEED 382
Db 365 MLELLDRVQGFRI DHE 381

RESULT 6
ADM97573
ID ADM97573 standard; protein; 403 AA.
AC ADM97573;
XX
XX DT 15-JUL-2004 (first entry)
XX
XX DE S ambofaciens spiramycin biosynthetic enzyme encoded by ORF1.
XX
XX KW antibacterial; immunosuppressive; antirheumatic; antiarthritic;
KW anthelmintic; insecticide; biosynthesis; macrolide; antibiotic;
KW spiramycin; veterinary medicine; immunosuppressant; transplant rejection;
KW rheumatoid arthritis; autoimmune disease; insecticide; anthelmintic;
KW avermectin.
XX
XX OS Streptomycetes ambofaciens.
XX
XX PN WO2004033689-A2.
XX
XX PD 22-APR-2004.
XX
XX PF 08-OCT-2003; 2003WO-FR002962.
XX
XX PR 08-OCT-2002; 2002FR-00012489.
XX 27-FEB-2003; 2003FR-00002439.
XX 07-AUG-2003; 2003US-0493490P.
XX
XX PA (AVET ) AVENTIS PHARMA SA.
XX (CNRS ) CNRS.
XX
XX PI Blondelet-Rouault M, Dominguez H, Darbon-Rongere E, Gerbaud C;
XX PI Gondran A, Karray F, Lacroix P, Oestreicher-Mermet- Bouvier N;
XX PI Pernodet J, Tuphile K;
XX
XX DR WPI: 2004-330455/30.
XX N-PSDB; ADM97572.
XX
```



```
XX 03-JUN-2004 (first entry)
DT ABR82126
XX ABR82126 standard; protein; 404 AA.
DE AC ABR82126;
XX 30-SEP-2003 (first entry)
KW DT Bacterial P450 enzyme SEQ ID NO:16.
XX DE Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
XX KW diol; alkene; chirality; thermotolerance; thermostability.
XX OS Eubacteria.
XX PN WO2003052050-A2.
XX PD 26-JUN-2003.
XX PF 05-AUG-2002; 2002WO-US024910.
XX PR 03-AUG-2001; 2001US-0309497P.
XX PA (DIVE-) DIVERSA CORP.
XX PI Weiner D, Burke M, Hitchman T, Pujol C, Richardson T, Short J;
XX WPI; 2003-541641/51.
XX DR N-PSDB; ACF06116.
XX PT Novel polypeptide having P450 enzyme activity and polynucleotides
XX encoding the polypeptide, useful for catalyzing the hydrolysis of
XX epoxides and arene oxides to their corresponding diols.
XX PS Claim 104; Page 58-59; 365pp; English.
XX CC The present sequence represents a bacterial P450 enzyme. P450 enzymes can
XX be used to catalyze the hydrolysis of epoxides and arene oxides to their
XX corresponding diols. P450 enzymes can also be used for hydrolysing an
XX alkene, for producing a compound of a desired chirality, and for
XX increasing thermotolerance or thermostability of a P450 polypeptide
XX Sequence 404 AA;
XX Query Match 20.7%; Score 430.5; DB 6; Length 404;
XX Best Local Similarity 28.8%; Pred. No. 2.2e-33;
XX Matches 114; Conservative 81; Mismatches 164; Indels 37; Gaps 11;
QY 24 SOLREEDPV-----HYESIDSYFISRYHDVRYILQHPDIFTTKSLVERABPMRGPVLA 78
DB 29 TELLEKEPVAIRLPFAARNEAWLVTRYEDVRAVTS DPR-FSRTALLQQVTKMGHVA 87
QY 79 QMHGKE-----HSAKRIVVRSFTGDALDHLSPLIKQNAENLLAPYLERGK-SOLVND 131
DB 88 SKAALNVADPPYHTQLRKAVTKAFTQSTRRLRPLAQAGTDRLLDAMEAAGRPA DLMKHL 147
QY 132 GKTFAVCVTMDMLGDKDEHEKISEHSGVADFTISISQSEARAHSLWCSEQLSOYLMP 191
DB 148 HGPLPMVAVCDLLGPIEDRAELASW-----POLTSSGPGPES-----SKAAQAIHGYVIR 200
QY 192 VKERRVNPBGDLTSLICTSEYEGMALSDKDLALILNVLLAATEPADKTLALMIYHLIN 251
DB 201 LLDRRAEPQDLAGVLAESLAEG-RITAEAVSLAMAILISGAHAVRNSANWVYLLT 259
QY 252 NPEQNDVLADRLSVRAIAETLRYKPPVQLI--PROLSQDPTVVGMEIKKDTIVFCMIG 309
DB 260 RPELADRLRAEPGLLPQAVDELLRWIPHRNGVGLPRIATEDVEVGGLIRAGEAVVASYL 319
QY 310 AANRDEPAFEQPDVFNTHREDLGKISAFSGAARHLAFSGIHNCVGAFAFAKNEIIVANI 369
DB 320 AANRDPAAAFEDPDLDFDREGIG-----HVSFGHPHCHMGAMLTRMSEVNLST 369
QY 370 VLDKMRNIRL---BEDFCYAESGLYTRGVSLLVAF 402
```

```
XX 03-JUN-2004 (first entry)
DT S. mycarofaciens midcamycin polyketide synthetase ORF19 protein.
DE KW macrolide; midcamycin biosynthesis; polyketide synthetase; PKS; enzyme;
XX antimicrobial; ORF19.
XX OS Streptomyces mycarofaciens.
XX PN JP2004049100-A.
XX PD 19-FEB-2004.
XX PF 19-JUL-2002; 2002JP-00210516.
XX PR 19-JUL-2002; 2002JP-00210516.
XX PA (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX WPI; 2004-174115/17.
XX DR N-PSDB; ADM45913.
XX PT New polynucleotides encoding proteins important in the biosynthesis of
XX midcamycin.
XX PS Claim 1; SEQ ID NO 20; 264pp; Japanese.
XX CC The invention relates to a novel polynucleotide, specifically a
XX biosynthesis gene, comprising a nucleotide sequence which encodes a
XX protein, where the gene is concerned with the biosynthesis of
XX midcamycin. The polynucleotide of the invention demonstrates
XX antimicrobial activity and may be useful for the biosynthesis of
XX midcamycin, as well as for manufacturing macrolide compounds other than
XX midcamycin. The current sequence is that of the Streptomyces
XX mycarofaciens midcamycin polyketide synthetase (PKS) ORF19 protein of
XX the invention.
XX Sequence 403 AA;
XX Query Match 20.8%; Score 432; DB 8; Length 403;
XX Best Local Similarity 30.0%; Pred. No. 1.6e-33;
XX Matches 111; Conservative 76; Mismatches 159; Indels 24; Gaps 6;
QY 22 YFSQLEEDPVHYESIDSYFISRYHDVRYILQHPDIFTTKSLVERABP-----VMRGP-V 76
DB 23 WFAFNRTHTHPVWDSRHAQVFRYDDYLTGNNPQFFSSDFNEVMPTPELEMVIGPQT 82
QY 77 LAQMHGKESAKRRIVVRSFTGDALDHLSPLIKQNAENLLAPYLERGKSDLVNDFGKTF 136
DB 83 IGALDPPAHGPKRLVSOAFTPRMARLEPRIRAVTQELLDVAVRQETIDVVGDLSYALP 142
QY 137 VCVTMDMLGLDKRDHEKISEW-----HSGVADFTISISQSEARAHSLWCSEQLSOYIM 190
DB 143 VIVIAELLGIPSGDRDVRFGWVDTLTLNGL-EYPNLPNFSSETIAPAL---KEMTDYLL 198
QY 191 PVIKERRVNPBGDLTSLICTSEYEGMALSDKDLALILNVLLAATEPADKTLALMIYHL 250
DB 199 HQIHAKREAPVDLLISGLVQAEQDKRLTDEIVNIVALLTAGHVSSSTLLSNLFLVLE 258
QY 251 NNPEQNDVLADRLSVRAIAETLRYKPPVQLIPROLSQDPTVVGMEIKKDTIVFCMIGA 310
DB 259 ENPQALADRLARELVTGAVETLRYRSPFNNIFRLKEDTDILGPKKGMVIAWSQS 318
QY 311 AANRDEPAFEQPDVFNTHREDLGKISAFSGAARHLAFSGIHNCVGAFAFAKNEIIVANI 370
DB 319 AANRDEPHEPPTDIFRR-----SSSRHMAFGIGIHHCLGAFARQSGKVVL EIM 369
QY 371 LDKMRNIRLE 380
DB 370 LDQVREFRID 379
```

Db 370 LLDYPRRLAGSAEDVWQSKGL-IRGKELLVTW 404
 RESULT 10
 AAB15502
 ID AAB15502 standard; protein; 396 AA.
 AC AAB15502;
 XX 14-FEB-2001 (first entry)
 XX Bacillus subtilis hydroxylating protein #2.
 XX DE
 XX Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
 KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
 XX OS
 XX Bacillus subtilis.
 OS XX
 PN WO200044886-A1.
 XX XX
 XX 03-AUG-2000.
 PD XX
 XX 28-JAN-2000; 2000WO-JP000472.
 PF XX
 XX 29-JAN-1999; 99JP-00021707.
 PR XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA XX
 XX Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
 PI WPI: 2000-548827/50.
 XX DR
 DR N-PSDB; AAA95667.
 XX PT
 XX New protein derived from Bacillus genus microorganism useful for
 PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
 PT inhibitors.
 XX PS
 PS Claim 10; Page 94-96; 111pp; Japanese.
 CC This sequence represents a novel Bacillus derived protein having the
 CC activity of producing a hydroxylated bicyclic compound or the
 CC corresponding lactone from a bicyclic compound or the corresponding
 CC lactone. The protein is used for preparing hydroxylated compounds useful
 CC as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and
 CC hypocholesterolaemic agents.
 XX SQ
 SQ Sequence 396 AA;
 Query Match 20.5%; Score 426.5; DB 3; Length 396;
 Best Local Similarity 26.7%; Pred. No. 5.4e-33;
 Matches 97; Conservative 86; Mismatches 153; Indels 27; Gaps 6;
 Qy 18 NPAYFSQLEEDPVHYESIDSFYTSRYHDVRYILQHPDIFTTKSLVERAEPVNRGPVL 77
 Db 25 HPFPWYESMRKQAPVDFSEENQVSWFLYDDVKVVGDKELFSS-CMPQQTSSI--GNSI 81
 Qy 78 AQMHGKHSKRRIVRSFIDGDLHLSPLIKQNAENLAPYLRGKSLVNDPQKTFV 137
 Db 82 INMDPPKTKIRSVNKAFTPRAMKQWEIOEITDELIQKQGRSEDFLVHDFSYPLV 141
 Qy 138 CVTMDMLGDKRDHEKISEHSGVADPITSISQSPPEARHSLW-----CSQLSQYLMP 191
 Db 142 IVISELGVPSAHEQFKAW----SOLLVSTPKDSEAEKAFLEERDKCEELAAFFAG 197
 Qy 192 VIKERRVNPQSDLSILCTSEVEGMALSDKILALILNVLNLAATBPADKTLALMTYHLIN 251
 Db 198 IIEKNKEQDIISILVEAEETGEKLSGELLPLCTLLLVAGNETTNILSNAMYSILE 257
 Qy 252 NPEQMDVLADRSVLPRAETLRYPKPVQLIPRQLSQDTPVVGGMIEKKOTIVFCMGAA 311
 Db 258 TPGVYSELASHPELMPQAEALRFRAPAPVLRRAKRTETIGGHLIKEGDMVLAFVASA 317
 Qy 312 NRDPFAEPQDPVNIHREDLGIKSAFSAARHLAEGSGIHNCVGAFAKNEIEIVANIVL 370
 Db 318 NRDEAKFDRPHMFDIKRH-----PNPIAFGHGHIHCIGAPLAKLE-----ANIAL 366
 Qy 372 DKM 374
 Db 364 TSL 366
 RESULT 11
 ABR82143
 ID ABR82143 standard; protein; 404 AA.
 AC ABR82143;
 XX 30-SEP-2003 (first entry)
 XX Bacterial P450 enzyme SEQ ID NO:50.
 XX Bacterial; P450; enzyme; catalysis; hydrolysis; epoxide; arene oxide;
 KW diol; alkene; chirality; thermotolerance; thermostability.
 XX OS
 OS Eubacteria.
 PN WO2003052050-A2.
 XX XX
 XX 26-JUN-2003.
 PD XX
 XX 05-AUG-2002; 2002WO-US024910.
 PF WPI: 2003-541641/51.
 XX DR
 DR N-PSDB; ACF06133.
 XX PT
 XX Novel polypeptide having P450 enzyme activity and polynucleotides
 PT encoding the polypeptide, useful for catalyzing the hydrolysis of
 PT epoxides and arene oxides to their corresponding diols.
 XX PS
 PS Claim 104; Page 360-361; 365pp; English.
 CC The present sequence represents a bacterial P450 enzyme. P450 enzymes can
 CC be used to catalyse the hydrolysis of epoxides and arene oxides to their
 CC corresponding diols. P450 enzymes can also be used for hydrolysing an
 CC alkene, for producing a compound of a desired chirality, and for
 CC increasing thermotolerance or thermostability of a P450 polypeptide
 XX SQ
 SQ Sequence 404 AA;
 Query Match 20.4%; Score 423.5; DB 6; Length 404;
 Best Local Similarity 26.3%; Pred. No. 1.1e-32;
 Matches 103; Conservative 89; Mismatches 160; Indels 39; Gaps 6;
 Qy 11 LSOQFQNNPAYFSQLEEDPVHYESID--SYFISRYHDVRYILQHPDIFTTK----- 62
 Db 3 ITPEFLRDPYPVYQVRRETGRMHLSANTGRWFLPHADIRALTARDERFSASRKGAGV 62
 Qy 63 -----SLVERAEPVNRGPVLAQMKGKHSKRRIVRSFIDGDLHLSPLIKQNAE 113
 Db 63 NQPPAEVRPEFARFNEAISRWIVLHQ--PEHQRLQLMQQGFTRLLITTMFKQRVCD 120
 Qy -114 NLLAPYLRGKSLVNDPQKTFVAVCTVMDMLGDKRDHEKISEHSGVADPITSISQSP 173
 Db 121 DLIDAFVGRGSTFMEYAHPPAKVIAEMLGWNPEDYPAFVWVSDLLNFAAGSLRPTLE 180
 Qy 174 ARAHSLWCSEQ-----LSQYLMPIVIRRVNPGVSDLSILCTSEVEGMALSDKILALIL 228
 Db 181 -----NFRAAQDQGLLAMDYFARLLPERENPGDGLVSLLSAEGEWMWTAEQVLANT 235

```
QY 229 NVLLAATEPADKTLALMIYHLLNPEQMNDVLADRSVPRAIAETLRYKPPVQLIPRQLS 288
Db 236 QIIVAGHETTRNLVANGVELLRYPEQRALLSRPELMPSAVREIMRFPESPLQFIRRVAR 295
QY 289 QTVVGGMEIKDITVFCMIGANRDPPEAFEPDVFNIHREDLGKSAFSGAARHLAFGS 348
Db 296 EDFEFGAEVREGDGLVLMGLSANRDPPEAFDDPTFDLTRNPTG-----HLAFGW 345
QY 349 GTHNCVGAFAKNEIEIVANIVLDKMNRL 379
Db 346 GHVCVGAALAELEGQVSPRTLLDRLPGLEL 376

RESULT 12
AAB15501
ID AAB15501 standard; protein; 396 AA.
XX
AC AAB15501;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating protein.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS Bacillus subtilis.
XX
PN WO200044886-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-JP000472.
XX
PR 29-JAN-1999; 99JP-00021707.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX
DR WPI; 2000-548827/50.
DR N-PSDB; AAA95666.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors.
XX
PS Claim 8; Page 68-71; 11pp; Japanese.
XX
CC This sequence represents a novel Bacillus derived protein having the
CC activity of producing a hydroxylated bicyclic compound or the
CC corresponding lactone from a bicyclic compound or the corresponding
CC lactone. The protein is used for preparing hydroxylated compounds useful
CC as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and
CC hypocholesterolaemic agents
XX
SQ Sequence 396 AA;

Query Match 20.1%; Score 418.5; DB 3; Length 396;
Best Local Similarity 26.2%; Pred. No. 3.3e-32;
Matches 95; Conservative 86; Mismatches 155; Indels 27; Gaps 6;

QY 18 NPYAYFSQLREEDPVHYESIDSYSFISYHVDVRYILOHPDIFTTKSLVERAEPVMRGPVL 77
Db 25 HFPFPMYEMRKADAPVSFDEENQWVSFLYDDVKVVGDKELFSS-CMPQQTSSI--GNSI 81
QY 78 AQMHGKHSKRIYVRSFTGDALDHLSPLIKQNAENLLAPYLERGKSLVNDFGKTFV 137
Db 82 INMDPPKHTKISVNVKATFTPRVMKQWEPRIOBITDELQKQGRSEFDLVHDFSYPLPV 141
QY 138 CVTMDMLGDKRDHEKISSEHWSGVADFTTISQSPARAHSW-----CSEQLSOYLM 191
Db 142 IVISELLGVPSAHMEQFKAW-----SDLLVSTPKDKSEAEKAFLEBRDKCEBELAAFFAG 197
```

```
QY 192 VIKERRVNPGLSILCTSEYEGMALSDKDILALILNVLLAANTEPADKTLALMIYHLIN 251
Db 198 IIEKRNKPEQDIISILVEAETGKLSGEBELIPFCTLLVAGNETTTNLSINAMYSILE 257
QY 252 NPEQNDVLADRSVPRAIAETLRYKPPVQLIPRQLSODTVVGGMEIKDITVFCMIGAA 311
Db 258 TPGVYELRSHPELMPQAVEALRPRAPVLRRIAKRDTTEIGGHLIKEGDMVLAFVASA 317
QY 312 NRDPFAEPQDVFNIHREDLGKSAFSGAARHLAFSGIHNVCVGAFAKNEIEIVANIVL 371
Db 318 NRDEAKFDRPHMFDIRRH-----PNPHIAPFGHGFCLGAPLARLE----ANIAL 363
QY 372 DKM 374
Db 364 TSL 366

RESULT 13
AAB15503
ID AAB15503 standard; protein; 396 AA.
XX
AC AAB15503;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating protein #3.
XX
KW Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
KW hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia.
XX
OS Bacillus subtilis.
XX
PN WO200044886-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-JP000472.
XX
PR 29-JAN-1999; 99JP-00021707.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX
DR WPI; 2000-548827/50.
DR N-PSDB; AAA95669.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors.
XX
PS Claim 10; Page 105-107; 11pp; Japanese.
XX
CC This sequence represents a novel Bacillus derived protein having the
CC activity of producing a hydroxylated bicyclic compound or the
CC corresponding lactone from a bicyclic compound or the corresponding
CC lactone. The protein is used for preparing hydroxylated compounds useful
CC as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and
CC hypocholesterolaemic agents
XX
SQ Sequence 396 AA;

Query Match 20.1%; Score 417.5; DB 3; Length 396;
Best Local Similarity 26.2%; Pred. No. 4.2e-32;
Matches 95; Conservative 87; Mismatches 154; Indels 27; Gaps 6;

QY 18 NPYAYFSQLREEDPVHYESIDSYSFISYHVDVRYILOHPDIFTTKSLVERAEPVMRGPVL 77
Db 25 HFPFPMYEMRKADAPVSFDEENQWVSFLYDDVKVVGDKELFSS-CMPQQTSSI--GNSI 81
QY 78 AQMHGKHSKRIYVRSFTGDALDHLSPLIKQNAENLLAPYLERGKSLVNDFGKTFV 137
```


Query Match 19.9%; Score 413; DB 6; Length 395;
Best Local Similarity 28.6%; Pred. No. 1.2e-31;

Matches	112;	Conservative	72;	Mismatches	162;	Indels	46;	Gaps	8;
Qy	8	FVSLSDQFQNPPYAFYSOLREEDPVHYEESIDSFYGRHVDVRYILQHDPDIFTT-----	61						
Dd	5	FNPYCERFYQPWFQNFRAIRLTDQDPVHIIEFDNALGFGEDEVWRAGMDRSFTATYGSSP	64						
Qy	62	-KSLVERAEAPVMRGPEVLQAOMHGKHSAKRRIVVRISFGIDALDLHLSPILKQNAENLLAPYL	120						
Dd	65	QALLIDR--VKQPEIFLFMDI PNMHIRGII AKDYGENAMPLEGGIKRATAKEAITYPL	121						
Qy	121	ERGKSDDLNDVFQKTFAVCVTWMDMLGLDKDRHEKI-----SEWHSGVADFITTSIQ	170						
Dd	122	KSGEMD-VYAFARIVALFTIADMIGLRPEEVVRIIRSLIDIFFGTTPGHRT-----	171						
Qy	171	SPEARASHLWCSEQLSOYLMPVIKERRVN--PGSDLISILCTSYEGVALSXDKDILAIL	228						
Dd	172	TPDGVA---AFHEVTAYVLDLI GHYRAKGAPEGSHIDNWLKAPDGPRLDDQALCANIF	227						
Qy	229	NVLAAATEPADKTLALMIYHLNNPQMNVDLAORSVPRAIAETLYRKPPVOLIPQLS	288						
Dd	228	SLSITGSDTVPLSSAAIYYLSEHPAQOLEAVRSDRALIIPAFAESTVRYDQPTNVLGRLLA	287						
Qy	289	QDTVVGMEI IKOTIVFCMIGGAANDPDEAPQPDVFNHRDELGIKSAFSGAARHLAFGS	348						
Dd	288	IDTDKYGKPMKKGOAVLFMYASANRDPLFEFHPDTFNIYRD-----PRRTLSPFGS	337						
Qy	349	GIHNCVGAFAFKNEIEIVANTIVDKMRNIRLE	380						
Dd	338	GIHICLGQLLAKLEGQIILETFPHIPDFTQQ	369						

Search completed: December 10, 2004, 13:40:37
Job time : 138.823 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:27:20 ; Search time 34.5732 Seconds
(without alignments)
776.868 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQSLFSLSDQFNPNP.....AESGLYTRGPVSLVAFDGA 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/1/1aa/5A-COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B-COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A-COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B-COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCUS-COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383.5	18.4	402	4	US-09-724-797-36
2	360	17.3	416	3	US-09-320-878-18
3	360	17.3	416	3	US-09-105-537-39
4	360	17.3	416	4	US-09-141-908-13
5	360	17.3	416	4	US-09-657-440-18
6	353.5	17.0	399	3	US-08-765-907A-10
7	352.5	16.9	419	3	US-09-335-409-8
8	352.5	16.9	419	3	US-09-413-814-71
9	352.5	16.9	419	3	US-09-568-102-8
10	352.5	16.9	419	3	US-09-567-969-8
11	352.5	16.9	419	3	US-09-568-486-8
12	352.5	16.9	419	3	US-09-568-486-8
13	352.5	16.9	419	3	US-09-568-472-8
14	352.5	16.9	419	3	US-09-567-899-8
15	347.5	16.7	406	6	5212296-6
16	335	16.1	403	6	5212296-9
17	335	16.1	448	4	US-09-724-797-44
18	302	14.5	437	4	US-09-252-991A-17836
19	273.5	13.1	412	1	US-08-102-863-11
20	273.5	13.1	412	5	PCR-US92-10885-11
21	271	13.0	409	3	US-09-385-028-12
22	271	13.0	409	4	US-09-726-614-12
23	271	13.0	409	4	US-09-385-040-12
24	257	12.4	468	4	US-09-252-991A-32437
25	249.5	12.0	395	4	US-09-266-965-129
26	238	11.4	189	4	US-09-679-279-20
27	221.5	10.6	504	3	US-09-126-420A-18

28	213	10.2	524	4	US-09-976-594-533	Sequence 533, Appl
29	197	9.5	444	3	US-09-413-814-95	Sequence 95, Appl
30	197	9.5	444	3	US-09-413-814-96	Sequence 96, Appl
31	195	9.4	414	4	US-10-018-730A-2	Sequence 2, Appl
32	195	9.4	509	4	US-10-140-002-108	Sequence 108, Appl
33	192.5	9.3	510	4	US-09-852-067-2	Sequence 2, Appl
34	192	9.2	513	4	US-09-142-108C-15	Sequence 15, Appl
35	192	9.2	526	1	US-08-298-426-4	Sequence 4, Appl
36	188	9.0	489	4	US-09-852-067-4	Sequence 23, Appl
37	186	8.9	476	4	US-09-142-108C-23	Sequence 4, Appl
38	184.5	8.9	484	4	US-10-142-231-56	Sequence 56, Appl
39	183	8.8	496	3	US-09-172-339-6	Sequence 6, Appl
40	182.5	8.8	503	4	US-09-583-447A-2	Sequence 2, Appl
41	182	8.8	496	3	US-09-292-768-64	Sequence 64, Appl
42	182	8.8	500	3	US-09-292-768-68	Sequence 68, Appl
43	178.5	8.6	422	1	US-08-396-218-2	Sequence 2, Appl
44	178.5	8.6	422	1	US-08-760-116-2	Sequence 2, Appl
45	178	8.6	496	3	US-08-881-784-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-724-797-36
; Sequence 36, Application US/09724797
; Patent No. 6733998
; GENERAL INFORMATION:
; APPLICANT: Jon S. THORSON
; TITLE OF INVENTION: MICROMONOSPORA ECHINOSPORA GENES
; TITLE OF INVENTION: ENCODING FOR BIOSYNTHESIS OF
; FILE REFERENCE: 2653-40
; CURRENT APPLICATION NUMBER: US/09/724,797
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 60/111,325
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Bacteria.
US-09-724-797-36

Query Match	18.4%;	Score 383.5;	DB 4;	Length 402;
Best Local Similarity	26.7%;	Pred. No. 1.6e-32;		
Matches 111;	Conservative 87;	Mismatches 180;	Indels 37;	Gaps 13;
Qy	4	SIKLFSLSDQFNPNPYAFSOLREEDPVHYEESIDSYFISRYHDVRYILOHPDIFTTKS	63	
Db	5	AVTAFDPTDADVRDRDPYSYHMLLRHDPVH-RGAHRVWVVSFADVRAVL--GDERFART	61	
Qy	64	LVERAEPVNRGP-VLAQMHG-----KEHSAKRIVVRISFGDLDHLSPLIKQNAEN	114	
Db	62	GIRRFWTDLVGPGLLAEIVGDIILFQDEPDHGLRGVGPAPSPALRELEPIACTVDD	121	
Qy	115	LIAPYLEKGSDLVNDFGKTFACVCTVMGLGDKRDHEKISEWHSGVADFI-TSISOSPE	173	
Db	122	LURPALARGAMDVVDELAYPELALRAVLGLLGLPADWAGVGRWSRDVGRTLDRGSAEDM	181	
Qy	174	ARAHSLWCSEQLSOYLMPIKERRVNPGLSILCTSEYEGMALSDKDTALILINLVLLA	233	
Db	182	RRGHA--ATAEPADYVERALARRRREGGBOLLALMLDAHDRGL-MGRNEIVSTVVTFT	238	
Qy	234	ATEPADKTLALMIYHLLNNPEQMDVLADR-SLVPRATAETLRYKPPVQLIPQLSODTV	292	
Db	239	GHEVTASQVGNVLSLLAHPDQL-DLLRRRDLQAQAVEECLRYDPSVQSNTRQLDDVDE	297	
Qy	293	VGMGEIKKOTIVFCMTIGANRDPPEAFQPDVFNHREDIGIKSAFSGAARHLAFSGSIHN	352	
Db	298	LRGRRLRRDDVVVVLAGAANRDRPRYDRPDFFDIERDPV-----PSMSFGAGMY	347	


```
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-141-908-13

Query Match
Best Local Similarity 17.3%; Score 360; DB 4; Length 416;
Matches 102; Conservative 76; Mismatches 173; Indels 30; Gaps 8;

QY 11 LSDQFQNNPYAFYFSQREEDPVH---YEEISIDSYFISRYHVDVRYLIQHP-----DIPTTKS 63
D 20 LQDFAADPYPTIYARLAEGPAHRVTPTEGDEVWLVVGYDRARAVLADPRFSKDWNRNSTT 79
QY 64 LVERAEPVMRGPVLAQMKGKHSKRRIIVRSFIFGDLHLSPLIKQNAENLLAPYLER- 122
D 80 PLTEAEEALNNHNL-ESDPPRTRLRKLVAREFTMRRELLRPVQEIVDGLVDAMLAAP 138
QY 123 -GKSDLVNDFGKTFVAVCTMDMLGDKRDHEKISEWHSGVADFITSISQSPPEARHSLWC 181
D 139 DGRADLMESLAWPLPITVISELGVPEPDRAAFRVW-----TDAFVFPDDPAQAQTA 190
QY 182 SEQLSOYLMMPVKERRVNPQSDLSILC-TSEYEGMALSDKDLILALINLVLLAATEPADK 240
D 191 MAEMGYLSRLSDSKRGQGEDLLSALVTSDEDSRLTSEELGMAHILLVAGHETTIVN 250
QY 241 TLALMIYHLNPNQMNVDLADSLVPRAIATLRYKPPVQLIPQLSQDTV-VGMEIK 299
D 251 LIANGMYALLSHPDQALALADMTLLDGAVEMLRYEGVESATYRFPVPEVDLQGTVP 310
QY 300 KDTIVFCMIGANRDEPEAQPDVFNHREDIGIKSAFSGAARHLAFSGIHCNCGAFA 359
D 311 AGDTVLVWLADAHRTPERPDPHREDIRDTAG-----HLAFGHIHFCIGAPLA 360
QY 360 KNEIEIVANIVLDKMRNIRLE 380
D 361 RLEARIARALLERCPLDALD 381

RESULT 6
US-08-765-907A-10
; Sequence 10, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE OF INVENTION: Mutasynthesis
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-10

Query Match
Best Local Similarity 17.0%; Score 353.5; DB 3; Length 399;
Matches 107; Conservative 65; Mismatches 175; Indels 63; Gaps 10;

QY 21 AYFSOLREEDPVHYBESIDSYFISRYHVDVRYLIQHPDIFTTK-SLVERAEPVMRGPVLAQ 79
D 18 AMLREWRHHHPVH-EDEYCAHFVFRHADVLTVASDPGVYSSQLSRRLRPGSQALSEQILSV 76
QY 80 MHGKHSKRRIIVRSFIFGDLHLSPLIKQNAENLLAPYLERGKSDLVNDFGKTFAVCV 139
D 77 IDPPMHRTLRRLVSQAFTPRTVADLEPRVTELQGLL-DAVDGDTFDLVAEPYLPVIV 135
QY 140 TMDMLGDKRDHEKISEW-----HSGVAD-----FITSISQSPPEARHSLWCS 182
D 136 IAEILGVPPADRTLFRSWSDRMLQMQVADPADMQFGDDADEDYQRLVLPKPFMRMHA---- 191
```

QY 183 EQLSOYLMPIKERRVWPGSDLSILCTSEYEGMALSDKDLALILNVLAAETPADKTL 242
Db 192 -----YLHDHVTDRARPANDLISALVAARVGERLTDEQIVFEGALLMAGHVTSMLL 246
QY 243 ALMIYHLNPNQMDVLDRLSLVPRALIAETLRYKPPVOLIIPQLSQDVTVVGMBIKKDT 302
Db 247 GNTVCLKDHPRAAARADRSILPALIEVLRPLPITVMARVTTKOTVLAGTIPAGR 306
QY 303 IVFCMIGANRDPEAFQPDVNIHREDLGKISAFSGAARHLAFSGIHCNCGAFAKNE 362
Db 307 MVVPSLLSANHQEVFTDPDHLDLARE-----GRQIAFGHGYICLGAFLARLE 355
QY 363 IRIIVANIVLDKMRNIRLEDFC-----YAESGLY-----TRGP 395
Db 356 GRALAEALFDRF-----PDFSPTDGAKURHYHRDGLFGVKNLPLTVRRGP 399

RESULT 7

US-09-335-409-8
; Sequence 8, Application US/09335409
; Patent No. 6121029

GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum

US-09-335-409-8

Query Match 16.9%; Score 352.5; DB 3; Length 419;
Best Local Similarity 27.2%; Pred. No. 3.8e-29;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSQFQNNPYAFYSQLEEDPVHYEESIDSYFISRYHDVRYILOHPDIFTTKSLVE- 66
Db 17 FKFPAPGYAEDPFAIERLEATPIFYWDEGRSWLTRYHDVSAVFRDRFAVSREEWES 76
QY 67 RAEPVMRGPVLAQM--HG-----KEHSKRRIIVRSFIGDALDHLSPKIKQNAENLLAP 118
Db 77 SAEYSSAIPELSDMKYGLGFLPPEDHARVRLVNPSTFSAIDLRLAEIQTVDQLLDA 136
QY 119 YLERGKSLVNDPFGKTFVAVCTVMDMLGDKRDHEKISEWHSGVA-----DFITSISQSP 173
Db 137 RSGQEEFVVRDYAEGIPMRALSALKVPAECDEKFRFGSATARALGVLVPQVDEETK 196
QY 174 ARAHSLWCSEQLSQYLMPIKERRVNP-GSDLSILCTSEYEGMALSDKDLALILNVL 232
Db 197 TLVASV--TEGLA-LLDVLDERRRNPENLVITMLQAEADGSLSTKELVALVGAIIA 253
QY 233 AATEPADKTLALMIYHLNPNQMDVLDRLSLVPRALIAETLRYKPPVOLIIPQLSQD 291
Db 254 AGTDTTIIYLIAPAVLNLLRSPLEALVKAEPGLMRNALDEVLRFDNLRIGTVRFAQDL 313
QY 292 VVGMEIKKDTIVFCMIGANRDPEAFQPDVNIHREDLGKISAFSGAARHLAFSGIHC 351
Db 314 EYCGASIKKGWVFLIPSLARDGTVFSRPDPVDFVRD-----TGAS--LAYGRGPH 363
QY 352 NCVGAAFAKNEIEIVANIVLDKMRNIRLEE 381
Db 364 VCPGVSLARLEAIEAVGTIFRRFPPEMKLKE 393

RESULT 8

US-09-413-814-71
; Sequence 71, Application US/09413814
; Patent No. 6225064

GENERAL INFORMATION:

; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum

US-09-413-814-71

Query Match 16.9%; Score 352.5; DB 3; Length 419;
Best Local Similarity 27.2%; Pred. No. 3.8e-29;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSQFQNNPYAFYSQLEEDPVHYEESIDSYFISRYHDVRYILOHPDIFTTKSLVE- 66
Db 17 FKFPAPGYAEDPFAIERLEATPIFYWDEGRSWLTRYHDVSAVFRDRFAVSREEWES 76
QY 67 RAEPVMRGPVLAQM--HG-----KEHSKRRIIVRSFIGDALDHLSPKIKQNAENLLAP 118
Db 77 SAEYSSAIPELSDMKYGLGFLPPEDHARVRLVNPSTFSAIDLRLAEIQTVDQLLDA 136
QY 119 YLERGKSLVNDPFGKTFVAVCTVMDMLGDKRDHEKISEWHSGVA-----DFITSISQSP 173
Db 137 RSGQEEFVVRDYAEGIPMRALSALKVPAECDEKFRFGSATARALGVLVPQVDEETK 196
QY 174 ARAHSLWCSEQLSQYLMPIKERRVNP-GSDLSILCTSEYEGMALSDKDLALILNVL 232
Db 197 TLVASV--TEGLA-LLDVLDERRRNPENLVITMLQAEADGSLSTKELVALVGAIIA 253
QY 233 AATEPADKTLALMIYHLNPNQMDVLDRLSLVPRALIAETLRYKPPVOLIIPQLSQD 291
Db 254 AGTDTTIIYLIAPAVLNLLRSPLEALVKAEPGLMRNALDEVLRFDNLRIGTVRFAQDL 313
QY 292 VVGMEIKKDTIVFCMIGANRDPEAFQPDVNIHREDLGKISAFSGAARHLAFSGIHC 351
Db 314 EYCGASIKKGWVFLIPSLARDGTVFSRPDPVDFVRD-----TGAS--LAYGRGPH 363
QY 352 NCVGAAFAKNEIEIVANIVLDKMRNIRLEE 381
Db 364 VCPGVSLARLEAIEAVGTIFRRFPPEMKLKE 393

RESULT 9

US-09-568-102-8
; Sequence 8, Application US/09568102
; Patent No. 6346404

GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan

```

; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; FILE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-8

Query Match
Best Local Similarity 16.9%; Score 352.5; DB 3; Length 419;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSQFQNNPYAFVSQIREEDPVHYEESIDSYFISRYHDVRYILOHDPDIETTKSLVE- 66
DB 17 FKFPAGYAEPPAERLREATPIFYWDEGSRWLTTRYHDSAVFRDRFPAVSREEWES 76
QY 67 RAEPVNRGVPVLAQM--HG-----KEHSKRRIVRSFSGDLDHLSPLIKNAENLLAP 118
DB 77 SAEYSSAIPELSDMKKYGLFGLPPEDHARVKLVNPSFTSRAIDLRLRAEIQTVDQLLDA 136
QY 119 YLERGKSDLVNDFGKTFVAVCVTMDMLGLDKRDHEKISEWHSGVA-----DFTISISQSPE 173
DB 137 RSGQEEFVVRDYAEGIPMRAISALLKVPACDEKFRFGSATARALGVLGPVQVDEETK 196
QY 174 ARAHSLWCSEQLSOYLMPIKERRVNP-GSDLSILCTSEYEGMALSDKDILALINVL 232
DB 77 SAEYSSAIPELSDMKKYGLFGLPPEDHARVKLVNPSFTSRAIDLRLRAEIQTVDQLLDA 136
QY 119 YLERGKSDLVNDFGKTFVAVCVTMDMLGLDKRDHEKISEWHSGVA-----DFTISISQSPE 173
DB 137 RSGQEEFVVRDYAEGIPMRAISALLKVPACDEKFRFGSATARALGVLGPVQVDEETK 196
QY 174 ARAHSLWCSEQLSOYLMPIKERRVNP-GSDLSILCTSEYEGMALSDKDILALINVL 232
DB 197 TLVASV--TEGLA--LLHDVLDERRNPLENDVITMLQAEADGSRSLSTKELVALVGAI 253
QY 233 AATEPADKTLALMIYHLNPNQMDVLDRLSPRAIAETLRYKPPVOL-IPRQLSQDT 291
DB 254 AGTDTTIVLIAFAVLNLRSPALSLVKAEPGLMRNALDEVLRFDNLRIGTVRFARQDL 313
QY 292 VVGGMEIKDITVFCMIGANRDPEAFQPDVFNHREDLGIKSAFSGAARHLAFSGIH 351
DB 314 EYCGASIKKGVNLLIPSALRDGTVFSRDPVDFVRRD-----TGAS--LAYGRGPH 363
QY 352 NCVGAFAKNEIETIVANIVLDKMRNIRLEE 381
DB 364 VCPGVSLARLEAEIAGVTIFRRFPPEMKLKE 393

RESULT 11
US-09-568-480-8
; Sequence 8, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-8

Query Match
Best Local Similarity 16.9%; Score 352.5; DB 3; Length 419;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

QY 8 FSVLSQFQNNPYAFVSQIREEDPVHYEESIDSYFISRYHDVRYILOHDPDIETTKSLVE- 66
DB 17 FKFPAGYAEPPAERLREATPIFYWDEGSRWLTTRYHDSAVFRDRFPAVSREEWES 76
QY 67 RAEPVNRGVPVLAQM--HG-----KEHSKRRIVRSFSGDLDHLSPLIKNAENLLAP 118
DB 77 SAEYSSAIPELSDMKKYGLFGLPPEDHARVKLVNPSFTSRAIDLRLRAEIQTVDQLLDA 136
QY 119 YLERGKSDLVNDFGKTFVAVCVTMDMLGLDKRDHEKISEWHSGVA-----DFTISISQSPE 173
DB 137 RSGQEEFVVRDYAEGIPMRAISALLKVPACDEKFRFGSATARALGVLGPVQVDEETK 196
QY 174 ARAHSLWCSEQLSOYLMPIKERRVNP-GSDLSILCTSEYEGMALSDKDILALINVL 232
DB 197 TLVASV--TEGLA--LLHDVLDERRNPLENDVITMLQAEADGSRSLSTKELVALVGAI 253
QY 233 AATEPADKTLALMIYHLNPNQMDVLDRLSPRAIAETLRYKPPVOL-IPRQLSQDT 291
DB 254 AGTDTTIVLIAFAVLNLRSPALSLVKAEPGLMRNALDEVLRFDNLRIGTVRFARQDL 313
QY 292 VVGGMEIKDITVFCMIGANRDPEAFQPDVFNHREDLGIKSAFSGAARHLAFSGIH 351
DB 314 EYCGASIKKGVNLLIPSALRDGTVFSRDPVDFVRRD-----TGAS--LAYGRGPH 363
QY 352 NCVGAFAKNEIETIVANIVLDKMRNIRLEE 381
DB 364 VCPGVSLARLEAEIAGVTIFRRFPPEMKLKE 393

RESULT 10
US-09-567-969-8
; Sequence 8, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
```


137 RSGQEFVVDVYAGGIPRAISALLKVPACDEKFRFGSATARALGVLVPQVDEETK 196
174 ARAHSLWCSEQLSQYLMPIKERRVNP-GSDLSILCTSEYEGMALSKDILAILNLVLL 232
197 TLVASV--TEGLA-LLHVDLDRRRNPLENDVLTMLLQAEADGSLSTKELVALVGAIIA 253
233 AATEPADKTLALMIYHLNNPEOMDNLADRSIVPRAIAETLRYKPPVQL-IPRQLSQDT 291
254 AGTDTTIIIAFVNLNLSPEALELVKAEPLMRNALDEVLFNLIIRIGTVRFARQDL 313
292 VVGMEIKKDTIVFCMIGANRDPFAEQDPVFNHREDLGIKSAPSGAARHLAFSGGTH 351
314 EYCGASIKKGEMVFLIPSALEDGTVFSRPDVFVDRD-----TGAS--LAYGRPH 363
352 NCVGAAFAKNEIEIVANIVLDKMRNIRLEE 381
364 VCPGVSLARLEAIEAVGTIFRRFPPEMKLKE 393

RESULT 12

US-09-568-486-8

; Sequence 8, Application US/09568486

; Patent No. 6355459

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,486

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 419

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-568-486-8

Query Match 16.9%; Score 352.5; DB 3; Length 419;
Best Local Similarity 27.2%; Pred. No. 3.8e-29;
Matches 106; Conservative 85; Mismatches 170; Indels 29; Gaps 10;

8 FSVLSQDFQNNPYAFVSQLEEDPVHYEESIDSYFISRYHDVRYILOHPDIPTTKSLVF- 66
17 FKPFAPGYAEDPFAIERLEATPIFYWDGSRSWLTRYHDVSAVFRDRFVSRREWES 76
67 RAEPVNRGPVLAQM--HG-----KEHSKRIRIVRSFTGDALDHLSPLIKQNAENLLAP 118
77 SAEYSSAIPELSDMKYKGLFGLPPEDHARVKLVNPSTFSTRADLLRAEIQRTVDQLDA 136
119 YLERGKSLVNDFGKTFVAVCVTMGLDKRDHEKISEWHSGVA-----DFITSISQSP 173
137 RSGQEFVVDVYAGGIPRAISALLKVPACDEKFRFGSATARALGVLVPQVDEETK 196
174 ARAHSLWCSEQLSQYLMPIKERRVNP-GSDLSILCTSEYEGMALSKDILAILNLVLL 232
197 TLVASV--TEGLA-LLHVDLDRRRNPLENDVLTMLLQAEADGSLSTKELVALVGAIIA 253
233 AATEPADKTLALMIYHLNNPEOMDNLADRSIVPRAIAETLRYKPPVQL-IPRQLSQDT 291
254 AGTDTTIIIAFVNLNLSPEALELVKAEPLMRNALDEVLFNLIIRIGTVRFARQDL 313
292 VVGMEIKKDTIVFCMIGANRDPFAEQDPVFNHREDLGIKSAPSGAARHLAFSGGTH 351
314 EYCGASIKKGEMVFLIPSALEDGTVFSRPDVFVDRD-----TGAS--LAYGRPH 363
352 NCVGAAFAKNEIEIVANIVLDKMRNIRLEE 381
364 VCPGVSLARLEAIEAVGTIFRRFPPEMKLKE 393

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:45:11 ; Search time 118.537 Seconds
(without alignments)
1220.359 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQSKLFSVLSDQFQNNPY.....AESGLYTRGPVSLVAFDGA 405

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1595576 seqs, 357178320 residues

Total number of hits satisfying chosen parameters: 1585576

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US04_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US03_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US02_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US01_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US00_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US10G_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US10H_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2080	100.0	405	15	US-10-627-124-2
2	454	21.8	410	14	Sequence 2, Appli
3	432	20.8	403	15	Sequence 6, Appli
4	430.5	20.7	404	14	Sequence 20, Appl
5	429.5	20.6	405	14	Sequence 16, Appl
6	423.5	20.4	404	14	Sequence 12073, A
7	413	19.9	395	14	Sequence 5073, A
8	412.5	19.8	408	14	Sequence 12, Appl
9	406	19.5	392	14	Sequence 4, Appli
10	399.5	19.2	425	14	Sequence 32, Appl
11	398	19.1	388	14	Sequence 20, Appl
12	396.5	19.1	399	14	Sequence 13776, A
13	392.5	18.9	400	15	Sequence 8126, Ap
					US-10-282-122A-62072

14	391	18.8	418	14	US-10-132-134-40	Sequence 40, Appl
15	389	18.7	393	14	US-10-156-761-9525	Sequence 9525, Ap
16	376	18.1	404	14	US-10-156-761-10431	Sequence 10431, A
17	374	18.0	430	14	US-10-145-415-95	Sequence 95, Appl
18	373.5	18.0	430	14	US-10-145-415-30	Sequence 30, Appl
19	371.5	17.9	430	14	US-10-145-415-4	Sequence 4, Appli
20	371.5	17.9	430	14	US-10-145-415-28	Sequence 28, Appl
21	370.5	17.8	457	14	US-10-156-761-11073	Sequence 11073, A
22	366.5	17.6	430	14	US-10-145-415-8	Sequence 8, Appli
23	366.5	17.6	430	14	US-10-145-415-12	Sequence 12, Appl
24	366.5	17.6	430	14	US-10-145-415-20	Sequence 20, Appl
25	365.5	17.6	404	14	US-10-214-446-40	Sequence 40, Appl
26	365	17.5	429	14	US-10-145-415-14	Sequence 14, Appl
27	362.5	17.4	424	15	US-10-229-148B-13	Sequence 13, Appl
28	361.5	17.4	404	14	US-10-321-188-76	Sequence 76, Appl
29	361.5	17.4	415	14	US-10-214-446-56	Sequence 56, Appl
30	361	17.4	414	14	US-10-156-761-13368	Sequence 13368, A
31	361	17.4	475	14	US-10-145-415-16	Sequence 16, Appl
32	360	17.3	416	9	US-09-861-289-39	Sequence 39, Appl
33	360	17.3	416	9	US-09-860-846-39	Sequence 39, Appl
34	360	17.3	416	10	US-09-988-384B-39	Sequence 39, Appl
35	360	17.3	416	10	US-09-836-821-39	Sequence 39, Appl
36	360	17.3	416	10	US-09-793-708-18	Sequence 18, Appl
37	360	17.3	416	14	US-10-201-365-13	Sequence 13, Appl
38	360	17.3	416	14	US-10-160-539-18	Sequence 18, Appl
39	360	17.3	416	14	US-10-271-889-39	Sequence 39, Appl
40	356.5	17.1	418	15	US-10-461-194-118	Sequence 118, App
41	356	17.1	475	14	US-10-145-415-22	Sequence 22, Appl
42	353	17.0	430	9	US-09-738-626-4117	Sequence 4117, Ap
43	352.5	16.9	399	14	US-10-314-657-37	Sequence 37, Appl
44	352.5	16.9	419	13	US-10-014-717-8	Sequence 8, Appli
45	352.5	16.9	430	14	US-10-145-415-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-10-627-124-2
; Sequence 2, Application US/10627124
; Publication No. US2004009694A1
; GENERAL INFORMATION:
; APPLICANT: NOVARTIS BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient Mutants Of Bacillus Cells
; FILE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-627-124-2

Query Match 100.0%; Score 2080; DB 15; Length 405;
Best Local Similarity 100.0%; Pred. No. 8.2e-187;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSQSKLFSVLSDQFQNNPYAFSOLREDDPVHYEESIDSYFISRYHDVRYILOHPDIFT	60
DB	1	MSQSKLFSVLSDQFQNNPYAFSOLREDDPVHYEESIDSYFISRYHDVRYILOHPDIFT	60
QY	61	TKSLVERAEPVNRGPGVLAQMKGKHSKRRIIVRSFGDLDHLSPLIKQNAENLLAPYL	120
DB	61	TKSLVERAEPVNRGPGVLAQMKGKHSKRRIIVRSFGDLDHLSPLIKQNAENLLAPYL	120
QY	121	ERKGSDLVNDFOKTFACVCTMDMLGDKRDHKKISEWHSGVADFTTISQSPEARASLW	180
DB	121	ERKGSDLVNDFOKTFACVCTMDMLGDKRDHKKISEWHSGVADFTTISQSPEARASLW	180

APP/CONF

QY 181 CSEQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLLAAATEPADK 240
DB 181 CSEQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLLAAATEPADK 240
QY 241 TLALMIYHLLNPNQNDVLADRSVLPRAIAETLRYKPPVQLIPQLSQDTVVGMEIKK 300
DB 241 TLALMIYHLLNPNQNDVLADRSVLPRAIAETLRYKPPVQLIPQLSQDTVVGMEIKK 300
QY 301 DTVFCMIGANRDPEAFQDPVFNHREDLGIKAFSGAARHLAFGSGIHCNCVGAFAK 360
DB 301 DTVFCMIGANRDPEAFQDPVFNHREDLGIKAFSGAARHLAFGSGIHCNCVGAFAK 360
QY 361 NEIEIVANVLDMKNIRLEEDFCVAESGLYTRGPVSLVAFDGA 405
DB 361 NEIEIVANVLDMKNIRLEEDFCVAESGLYTRGPVSLVAFDGA 405

RESULT 2
US-10-205-032-6
; Sequence 6, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 410
; TYPE: PRT
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-6

Query Match 21.8%; Score 454; DB 14; Length 410;
Best Local Similarity 29.3%; Pred. No. 9.8e-34;
Matches 117; Conservative 84; Mismatches 156; Indels 42; Gaps 9;
QY 22 YFSQLEEDPVHYEBSIDSYFISRYHDVRIYILQHPDI-----FTTKSLVERABPVMRGP-V 76
DB 26 WFAYMRKNWPVSWDETRQAWHVFSDYQVTNPLIFSSDFTSVFPVPSLALLMGPT 85
QY 77 LAQMHGKHSKRIRVRSFIDGDLHLSPLIKQNAENLLAPYLRGKSDLVNDFGKTF 136
DB 86 IGGIDPPRHAPRLKLVQAFTPRRIAQMELRIGQITADVLDQVRDQDRIDIASDLAYPLP 145
QY 137 VCVTMDMLGDKRDHEKISEW-----HSGVA-----DFITSISQSPEARAHSLWCSEQ 184
DB 146 VTVIAELGIPDKHEKFEWVDIILNSGLEVPNLPDDFTETVGFA-----IEE 195
QY 185 LSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLLAAATEPADKTLAL 244
DB 196 WSEFLYAQIAHKAEPKODLISGLCAAEDVGKLTDEEVNIVA-LLLTAGHISATLLS 254
QY 245 MYHLL-NNPEQNDVLADRSVLPRAIAETLRYKPPVQLIPQLSQDTVVGMEIKKXOTI 303
DB 255 NLFVLEEHQQAQAAVRADRSVLPVGIETLRYSPFNCIFRILNEDTDLGHMPRKGM 314
QY 304 VFCMIGANRDPEAFQDPVFNHREDLGIKAFSGAARHLAFGSGIHCNCVGAFAKQEI 363
DB 315 VTAIASANRDTEVFDPTDFIRRE-----SNKHLAFGHIHCLGAFALARUEA 364
QY 364 EIVANIVLDMKNIRLEE---DFCAESGLYTRGPVSL 399
DB 365 KVFLNQTLDOQTEFRIDHVGVEFYDADQLTARLFPQVV 403

RESULT 3
US-10-229-148B-20

; Sequence 20, Application US/10229148B
; Publication No. US20040091975A1
; GENERAL INFORMATION:
; APPLICANT: Meiji Seika Kaisha, Ltd.
; TITLE OF INVENTION: Midecamycin biosynthetic genes
; FILE REFERENCE: 138451 US
; CURRENT APPLICATION NUMBER: US/10/229,148B
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Streptomyces mycarofaciens
US-10-229-148B-20

Query Match 20.8%; Score 432; DB 15; Length 403;
Best Local Similarity 30.0%; Pred. No. 1.1e-31;
Matches 111; Conservative 76; Mismatches 159; Indels 24; Gaps 6;
QY 22 YFSQLEEDPVHYEBSIDSYFISRYHDVRIYILQHPDIFTTKSLVERAEP-----VMRGP-V 76
DB 23 WFAFNRTTHPVFWDESRHAWQVFRYDDYLTVSNNPQFFSSDFNEVMPTPELEWVGPT 82
QY 77 LAQMHGKHSKRIRVRSFIDGDLHLSPLIKQNAENLLAPYLRGKSDLVNDFGKTF 136
DB 83 IGALDPPAHGPMRKLVSQAFTPRRVARLEPRIRAVTQELLDVARGOETDVTVDGLSYALP 142
QY 137 VCVTMDMLGDKRDHEKISEW-----HSGVADFITSISQSPEARAHSLWCSEQSLYL 190
DB 143 VIVIAELGIPSGDRDVRGWDVTLITNEGL-EYNLPDNFSETTAPAL---KEMTDYLL 198
QY 191 PVIKERRVNPVSGDLISILCTSEYEGMALSDKDLALILNVLLAAATEPADKTLALMIYHLL 250
DB 199 HQIHAKREAPVDDLLISGLVQAEDQGRKLTDEIVNIVALLTAGHVSSSTLLSNLFLVLE 258
QY 251 NNPEQNDVLADRSVLPRAIAETLRYKPPVQLIPQLSQDTVVGMEIKKDTIVFCMICA 310
DB 259 ENPQALADLRADRELVTVGAVEETLRYSPFNFRFLKEDTDLILGPEMKKGQVIAWSQS 318
QY 311 ANRDPEAFQDPVFNHREDLGIKAFSGAARHLAFGSGIHCNCVGAFAKNEIEIVANIV 370
DB 319 ANRDPEHPEPDTDIR-----SSSSEHMAFGIIGHCLGAFALARQEGKVVLLELM 369
QY 371 LDKMRNIRLE 380
DB 370 LDQVREFRID 379

RESULT 4
US-10-214-446-16
; Sequence 16, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 404

[illegible][illegible]

QY 289 QDTVVGMEIKDITVFCMIGANRDPEAFQPDVFNHREDLGKSAFSGAARHLAFGS 348
Db 296 EDFEGGAELVRGDLVLMLGSANRDPEAFDPTDFDLTRNFTG-----HLAFGW 345
QY 349 GHNCVGAFAKNEIEIVANIVLDKMRNRL 379
Db 346 GPHVCVGAALAEQGVSRFTLLDRLPGLEL 376

RESULT 7

US-10-214-446-12
; Sequence 12, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-214-446-12

Query Match 19.9%; Score 413; DB 14; Length 395;
Best Local Similarity 28.6%; Pred. No. 6.7e-30;
Matches 112; Conservative 72; Mismatches 162; Indels 46; Gaps 8;
QY 8 FSVLSDQFQNNPVAYFSQLEEDPVHYEESIDSYFSRYHDVRYILQHPDIFT-----61
Db 5 FNPYCEEFTQNPQNFALRTQDPVHYIEFDNALFGFEDVWRAGMDRESTATYGGSP 64
QY 62 -KSLVERAEPVVRGQVLAQMHGKHSKRRIIVRSFIGNDALDHLSPLIKQNAENLLAPYL 120
Db 65 QALLIDR---VKQPEIFLFWMDIPNHHIRGIIAKDYGRNAMPLEKIRATAKEAITPYL 121
QY 121 ERGKSLVNDPDKTFAVCVTMDMLGDKRDHEKI-----SEWHSQVADFTTISQ 170
Db 122 KSGEMD-VYAFARTVALFTIADMIGLRPEEVVRIRSLIDIFFGRTPGHRT-----171
QY 171 SPEARAHSLWCSEQLSQYLMPIVKERRVN--PGSDILISILCTSEYEGMALSDKDILALIL 228
Db 172 TPDGVA-----AFHEVAYVLDLIGHYRAKAGPEGSHIDNWLKAEPPGRPLDQALCANIF 227
QY 229 NVLLAATEPADKTLALMIYHLNPNQMDVLDRLSLVPRALAEITLYRKPVPQLIPRQUS 288
Db 228 SLSITGSDTVPLSSAAIYYLSEHPAQLEAVRSRALIPAAFAETVRYDQPTNVLGRLLA 287
QY 289 QDTVVGMEIKDITVFCMIGANRDPEAFQPDVFNHREDLGKSAFSGAARHLAFGS 348
Db 288 IDTDKYGKPMKQQAQLFMYASANRDLPEFHPDFTENIYRD-----PRRTLSFGS 337
QY 349 GHNCVGAFAKNEIEIVANIVLDKMRNRL 380
Db 338 GHICLQGLLAKLEGGQILLETTFEHIPTDFTVQ 369

RESULT 8

US-10-214-446-4

; Sequence 4, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-4
Query Match 19.8%; Score 412.5; DB 14; Length 408;
Best Local Similarity 26.6%; Pred. No. 7.8e-30;
Matches 109; Conservative 88; Mismatches 166; Indels 47; Gaps 10;
QY 16 QNNPYAYFSQLEEE---DPVHYEESIDSYFSRYHDVRYILQHPDIFTTKSLV-----65
Db 19 QNPYPFLFERIREHGQVRQLNPTLEVMMVTGYDEAVAALTDPRLSSSPVGVNGLEEM 78
QY 66 ---ERAEPVVRGQVLAQMHGKHSKRRIIVRSFIGNDALDHLSPLIKQNAENLLAPYLER 122
Db 79 AHQERTNVLMSMLVA--NGEDHTLRNLVSKAFTRARVEQLAPRQVQAHDTDAFLDAVAAR 136
QY 123 GKSLVNDPDKTFAVCVTMDMLGDKRDHEKISEWHSQVADP-----ITSISQSPEARA 176
Db 137 GSADLVSEFALPLPNAVLSDLIGIPA-----EGQDPFARLAVGLIMPNTPERLA 186
QY 177 HSLWCSEQLSQYLMPIVKERRVNPQSDILISILCTSEYEGMALSDKDILALINVLAAATE 236
Db 187 KGARARAEITFEFFELIAQKKEPKDILLSALCAAQAE-ERISDRELTANTILLTAGHE 245
QY 237 PADKTLALMIYHLNPNQMDVLDRLSLVPRALAEITLYRKPVPQL-IPRQLSQDITVVG 295
Db 246 TTASLIANGVHALLRHPQFATLRDDPSLLPGAIEELLRYEGFVSRGVARFTTDPYEIG 305
QY 296 MEIKDITVFCMIGANRDPEAFQPDVFNHREDLGKSAFSGAARHLAFSGIHNCVG 355
Db 306 VTVPAGEMIITGLAANRDPAFYDRPDILDVAREV-----PQQLAFGHGVHFCIG 356
QY 356 AAFAPKNEIEIVANIVLDKMRNRL---BEDFCYAESGLYTRGFVSLVAF 402
Db 357 APLARAARIAIGTLRRFPDLRLADPDADLSRREGIL--RGWATLPVTF 404
RESULT 9
US-10-214-446-32
; Sequence 32, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05


```
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Bacterial
; US-10-214-446-32

Query Match
Best Local Similarity 19.5%; Score 406; DB 14; Length 392;
Matches 113; Conservative 65; Mismatches 172; Indels 26; Gaps 8;

QY 15 FQNNPYAYFSQREEDPVHYEESIDSY---FISRYHDVRYIQLQHPDIFTTKSL--VERAE 69
DB 13 FTANPPYVYAKLREAGVHEVRMPDGFQFWLVVVGHEGRAALADPRLAKSPSIVGVRPE 72
QY 70 PVMRGVPLAQMKGKHSKRRIIVRSFIGDALDHLSPLIKONAEINLLAPYLRGKSDLYN 129
DB 73 EDIIGVHLLAADAPDHTLRRLRVGTGTRRVEGLRPRIQQLTTELADAMEPAGRADLYD 132
QY 130 DFGKTFVAVCVTMDMLGDKRDHEKISEHSGVADFTISQSPEARAHSLWCSEQLSQYL 189
DB 133 AFAYPLPIVICELGVPADRDFTFRNSN---QLVPTGDOEFGQA-----MVDFAAYL 184
QY 190 MPVIKERR-VNPGSDIISILCTSEYV-GMALSDKOILALINVLAAATBPADKTLALMTY 247
DB 185 DALIEDKRAAGPTDOLLSALITARAEDGRLSGPELRANAYLLIAGHETTNNLIANTVR 244
QY 248 HLLNPEQNDVLAARSLVPRAIAETRLRYKPPVQLIPROLSDQTV-VGMEIKKOTIVFC 306
DB 245 NLLTHPEQLAALRADPDLLDGTTEESLYDGPVETGTFRTREAVTIGGREIAAGQYVLV 304
QY 307 MTGAARNDPEAEQPDVFNHREDLGKSAFSGAARHLAGSGIHCNVCVGAFAKNEIEIV 366
DB 305 GIGALDRDPARPPDPDFDIRDRTRG-----HLAFGHHYCLGAPLARLEGRIA 354
QY 367 ANIVLDKMNIRLEED 382
DB 355 LRTLLDRFPDLELDPE 370

RESULT 10
US-10-214-446-20
; Sequence 20, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Bacterial
; US-10-214-446-20

Query Match
Best Local Similarity 19.2%; Score 399.5; DB 14; Length 425;
Matches 113; Conservative 92; Mismatches 163; Indels 69; Gaps 12;

; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Bacterial
; US-10-214-446-32

Query Match
Best Local Similarity 30.1%; Pred. No. 3e-29;
Matches 113; Conservative 65; Mismatches 172; Indels 26; Gaps 8;

QY 14 QFQNNPYAYFSQREEDPVH---YBESIDSYFISRYHDVRYIQLQHPDIFTTKSLVERAE- 69
DB 23 EFASDPYPAYAWLRSHAPVHRTTLPSPGVAEWLVTTRYGDARQAL--ADQRLSKNPAHDES 80
QY 70 PVMRGVPLAQMKGKHSKRRIIVRSFIGDALDHLSPLIKONAEINLL 116
DB 81 PHAKGTGIPGERKAEMLTHLLNIDPPDHTLRRLRVSKAFTPRVAEFTPRVQOELTDRLI 140
QY 117 APYLRGKSDLYNDFCKTFVAVCVTMDMLGDKRDHEKISEW-----HSGVADFI 165
DB 141 DAFVTKGSADLIHDAFPLPIVAICDLGLVPEEDQDDFRDWAGMMIRHGGGPRGVSARV 200
QY 166 TSISQSPPEARAHSLWCSEQLSOYLMPVIKERRVNPVGSOLISILCTSEYEGMALSDKILA 225
DB 201 -----KMRGYLAELIHRKEAPGGDLISGLIKASDRGHEHLTENEAAA 243
QY 226 LILNVLAAATEPADKTLALMIYHLLNPP---BOMNDVLA--DRSLVPRAIAETRLRYKPPV 280
DB 244 MAFILLFAGFTVNLGNGVYQLLRHPQORRLQTSLAAGETGLLETGIEELLRYDGPV 303
QY 281 QLIP-RQLSQDVTVVGMEIKKDTIVFCMTGAANRDPEAEQPDVFNHREDLGKISAFSG 339
DB 304 EMATWRYATEPLTIGQODIPAGDPVLVLAADRDPERFDRPDVLDLARRD----- 354
QY 340 AARHLAFSGIHCNVCVGAFAKNEIEIVANIVLDKMNIRLEEDFCYAE-----SGLYTRGP 395
DB 355 -NQHLYGHHYCLGAPLARLEGOTATLTLTRLPDLRLAAD--PAELWRGGLIMRGL 411
QY 396 VSLVAVF 402
DB 412 RTLPEVF 418

RESULT 11
US-10-156-761-13776
; Sequence 13776, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13776
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-13776

Query Match
Best Local Similarity 19.1%; Score 398; DB 14; Length 388;
Matches 110; Conservative 70; Mismatches 189; Indels 26; Gaps 7;

QY 18 NPYAYFSQREEDPVHYEESID---SYFISRYHDVRYIQLQHPDIFTTK-----SLVERA 68
DB 10 DPAVYVYDLRDTAPVHRIAGTDGKPAWLVYDDVREGLANPULSLDKKHALPGNRYGLA 69
QY 69 EPVMRGPVLAQMKGKHSKRRIIVRSFIGDALDHLSPLIKONAEINLLAPYLRGKSDLY 128
DB 70 LPPALDANLNNWDAPDHTLRRLRVGTGTRRVEGLRPRIQQLTTELADALGTHGSTDLI 129
QY 129 NDFGKTFVAVCVTMDMLGDKRDHEKISEHSGVADFTISQSPEARAHSLWCSEQLSQY 188
```

Db 130 ASYAAPLPITVICDLGVPDEHRRDFRAW---TDPLVTPDPARPDPVARESV---VSLIGF 183
QY 189 LMPVTKERRVNGSGLISILCTSEYEGMALSOKDILALINVLAAATEPADTKTALMIYH 248
Db 184 FTGLADKRKNPADLLSLIAVQBEGDRLTDELMSLAFLILFAGYENTVHLIGNAVLA 243
QY 249 LANNPQONNDVLADSLVPRAETLRYKPPVQLIPRQLS-ODTVVGGMEIKKOTIVFCM 307
Db 244 LURHEQALAEAREDPARLPDAVEFARYEGPALLAIRPPVRDVTGIGTVVPAGETVLLS 303
QY 308 IGAANRDPFAFPQPDVFNTHREDLGIKSAFSGAARHLAFSGIHNCVGAAPAKNEIEIVA 367
Db 304 LSAANRDPFRFPDP-----RLDLG-----RDAAGHLALGHGVHCLGAPLARLETEVAL 353
QY 368 NIVLDMNRIRLEEDFCYAESGLYTRGVPVSLIVAF 402
Db 354 AALLERFPDLAAETEPRRPSLRARGLLALPVTY 388

RESULT 12

US-10-156-761-8126
; Sequence 8126, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8126
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8126

Query Match 19.1%; Score 396.5; DB 14; Length 399;
Best Local Similarity 28.4%; Pred. No. 2.4e-28;
Matches 111; Conservative 62; Mismatches 189; Indels 29; Gaps 7;
QY 18 NPAYVFSQLREEDPVHYEESIDSYFISRYHVRVYILQHPDIFTTKSLVERAEP-----VNR 73
Db 18 DPYPLVLAELREAGPV-VQADGSYLIGTYHEIVALLHDPMSADPR--SRTAPAPVEVTR 74
QY 74 GPVLAQMHGKEHSKRIRVVRFP-----IGDALDHLSPLIKQNAENLLAPYLERGKSDLVN 129
Db 75 KPSFLRLDDPEHRLNSAMRPFPHSPGRVDSMRGEIVQLTKELAEVFEQGRQIDVDV 134
QY 130 DFGKTFVAVCTVMDLGLDKRDEKISEHSGVADFTTISQSPPEARAHSLS-----WCSEQ 184
Db 135 DFAYPLPVTVICLLGIPKDKQLQFDW---TDTLVASADIGPEGDTAERDQAQAQOE 191
QY 185 LSQYLMPIVKERRVNGSGLISILCTSEYEGMALSOKDILALINVLAAATEPADTKTAL 244
Db 192 MGQYLVQLAEQRRGRPTGMDLSDLVNPEOPAAARLSEEDLAANTILLFIAGHETTNNLIAN 251
QY 245 MYHLLANPEQONNDVLADSLVPRAETLRYKPPVQLIPRQLSODTVVGGMEIKKOTIV 304
Db 252 GVLTLRLRPDQDRUREDSLLPRAVEELLRYEPPVHMRERVPLDIDVAGTTIPGTSV 311
QY 305 FCWIGAANDRDPFAEQPDVFNTHREDLGIKSAFSGAARHLAFSGIHNCVGAAPAKNEIE 364

Db 312 ILALASGSRDPMRFSEPDPRDFOTRPD-----NQHVGFSGIHLFCGAPLARIEAE 361
QY 365 IVANIVLDMNRIRLEEDFCYAESGLYTRGP 395
Db 362 AALGALLPHLGLTARLVQDPPPPYRQNAMLRGP 392
RESULT 13
US-10-282-122A-62072
; Sequence 62072, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62072
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-10-282-122A-62072

Query Match 18.9%; Score 392.5; DB 15; Length 400;
Best Local Similarity 26.9%; Pred. No. 5.8e-28;
Matches 108; Conservative 84; Mismatches 173; Indels 37; Gaps 8;
QY 18 NPAYVFSQLREEDPV-----HYEESIDS-----YFISRYHVRVYILQHPDIFTTKSLV 65
Db 18 DPYPLFARRRAGVAGTVMDSKTPKQEYSAVSFADAVNTVFRDGRVFSKPY 76
QY 66 ERAEPVMRGPVLQAMHGKHSKRIRVVRSPFTGCDALDHLSP-LIKQNAENLLAPYLERGK 124
Db 77 DKTIGLFWGPTTLANEGKKGRDHRNLVSAAFKSKALARWEPTIVRPICNALIDDFDAGT 136
QY 125 LVNDFGKTFVAVCTVMDLGLDKRDEKISEHSGVADFTTISQSPPEARAHSLSWC 191
Db 137 ADLVQRQTFEFPTRVIRALLGLPDED---LPMFHTRAVOLISYHVDYERAFEA----- 186

QY 182 SEOLSOYLMPEVTKERRVNPQSGDLISILCTSEYEGMALSKDILALINVLNLAATEPADKT 241
Db 187 SAALKDYFLEQIEQRKSKPTEDIGLVTAETIDGKLSDEAIYSFLRLPAGLETTERS 246
QY 242 LALMIYHLNPNQMDVLADSLVPRATAETLYRKPVPQVLIQVLSQDVTVVGMEIKD 301
Db 247 SGNLLYLLTHPDQFALQADRELLAPALEGURFETPLTTVORFTTEDEHLGVRIAPAR 306
QY 302 TIVFCMIGANRDPEAFQPDVFNHREDLGKISAFSGAARHLAFSGIHNCVGAFAKN 361
Db 307 SVIGVCIGSANDRERWSEDFIKHV-----PHISPAAGHTCLGLHLARL 356
QY 362 EIEIVANIVLDKRNIRLEEDFCYAESGLYTRGVSLLVAFD 403
Db 357 ETRVAMECLNRLTNVLLSDGPHIHGQFRSPALPVTFD 398

RESULT 14

US-10-132-134-40
; Sequence 40, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 40
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Streptomyces amphibiosporus
US-10-132-134-40

Query Match 18.8%; Score 391; DB 14; Length 418;
Best Local Similarity 26.7%; Pred. No. 8.6e-28;
Matches 111; Conservative .84; Mismatches 166; Indels 54; Gaps 11;
QY 10 VLSDQFN-----NPYAYFSQUREEDPVHYESIDSYFISGRVHDVYILOHPDFTTKSL 64
Db 17 IVTDVFPQSTEGKKNPYRLYRQLQELQVHRSEQL-GWVATGYEVCSAALRDPRIKGPQ 75
QY 65 VE-----RAEPVWRGCVLAQMHGKHSKRRIWVRSFIGDALDHLSPLIKQNAENL 115
Db 76 IQGPRDPAHSAEALLRG-TWRLDPPDHTLRRLVNGAFTPRSVAALPEPDIQELIDDL 134
QY 116 LAPYLERGKS----DLVNDFGKTFACVCTVMDMLGDKRDKHEKISEHWSGVADFITSI--- 168
Db 135 ITPAVKKAEGAPVDMMSGFAPLSVAVIGMLGVPSADWHR----PHDVVLDSNVELG 191
QY 169 ----SOSPEARHSLWCSEQLSOYLMPEVTKERRVNPQSGDLISILCTSEYEGMALSKDILA 225
Db 192 FTGDELPKADA---AADELIAYFRKLGAEERNPADDLTSTLANATEAGDRLTEQELVT 247
QY 226 LILNVLNLAATEPADKTLALMIYHLNPNQMDVLADSLVPRATAETLYRKPVPQVLI--- 283
Db 248 MLILFLMAGFTTTTHSMGMNGFALLNPEQTLWRNMDAMPAAVEELIRYSPVQFIAG 307
QY 284 ----PROLSQDVTVVGMEIKKDTIVFCMIGANRDPEAFQPDVFNHREDLGKISAFSG 339
Db 308 YTKPEVELADGTAV-----PADEYLFMLMIGANRDPRVSDPELLRLDR-----G 352
QY 340 AARHLAFSGIHNCVGAFAKNAEIEIVANIVLDKRNIRLEEDFCYAESGLYTRG 394
Db 353 EAAPMSFGGHIHYCLGAGLARLEIRKIPTSLTRFSAILARPEPERRSGLALRG 407

RESULT 15

US-10-156-761-9525

; Sequence 9525, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9525
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9525

Query Match 18.7%; Score 389; DB 14; Length 393;
Best Local Similarity 27.1%; Pred. No. 1.2e-27;
Matches 106; Conservative 79; Mismatches 178; Indels 28; Gaps 7;
QY 14 QFQNNPYAYFSQUREEDPVHY-----ERSIDSYFISGRVHDVYILOHPDIFT--TKSLVE 66
Db 13 RFTEDPYVYAELERRERGPVHWVTPPEAFEGMLVVGHEEAAALADPRLSKDGTKKGLT 72
QY 67 RAEPVWRGCVLAQMHGKHSKRRIWVRSFIGDALDHLSPLIKQNAENLAPVLERGKSD 126
Db 73 SLDVDLMGFIYLVNDPPETRLRSIVARAFTWRVREALRPRIQETDGLDMLPRGRAD 132
QY 127 LVNDFGKTFACVCTVMDMLGDKRDKHEKISEHWSGVADFITSIQSPEARHSLWCSEQLS 186
Db 133 LVDSFAYPLPITVICELGVPDIDRVTFRLSNEIV-----APTGGDAELAAVERLA 184
QY 187 QYLMPEVTKERR-VNPGSDLI-SILCTSEYEGMALSKDILALINVLNLAATEPADKTLAL 244
Db 185 AYLDLIDDKRSTAPADLLGLDIRABDDDLRLSGEELRAMAFILLVAGHETTNNLITN 244
QY 245 MIYHLNPNQMDVLADSLVPRATAETLYRKPVPQVLIQVLSQDVTVVGMEIKKDTI 303
Db 245 GVHTLLTHPDQFALQADRELLAPALEGURFETPLTTVORFTTEDEHLGVRIAPAR 304
QY 304 VFCMIGANRDPEAFQPDVFNHREDLGKISAFSGAARHLAFSGIHNCVGAFAKNAE 363
Db 305 VMIGLDAAGRDPAHPDPHVFDIHRAPQG-----HLAFGHIHYCLGAPLARLEA 354
QY 364 EIVANIVLDKRNIRLEEDFCYAESGLYTRG 394
Db 355 RVALSLLERCDDLALDGGPPGARPPGMLIRG 385

Search completed: December 10, 2004, 14:02:29
Job time : 123.537 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 11:14:31 ; Search time 417.058 Seconds
(without alignments)
9477.854 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagagatgacgggaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	753	100.0	753	12	ADJ62044
2	61.6	8.2	705	10	ACF70114
3	61.6	8.2	110000	10	ACF65385_2
4	61.6	8.2	110000	10	ACF67367_31
5	58.4	7.8	651	10	ACF66408
6	40.4	5.4	2000	8	ADA71938
7	39.8	5.3	2000	8	ADA71938
8	38	5.0	1194	8	ACA34459
9	38	5.0	110000	2	AAT42063_15
10	36.8	4.9	110000	10	ADH10017_0
11	36.8	4.9	110000	10	ADH10017_1
12	36.4	4.8	3048	2	AAQ75343
13	36.4	4.8	3211	2	AAQ75344
14	36.2	4.8	110000	6	ABA03041_24
15	36	4.8	1338	10	ACF72025
16	36	4.8	5580	4	ABL15162
17	36	4.8	110000	10	ACF67367_52
18	36	4.8	110000	10	ACF65387_3
19	35.4	4.7	736	4	AAH99421
20	35.4	4.7	825	4	ABL04657
21	35.4	4.7	1695	4	ABL20607

ALIGNMENTS

RESULT 1
ADJ62044
ID ADJ62044 standard; DNA; 753 BP.
XX
AC ADJ62044;
DT 06-MAY-2004 (first entry)
XX
DE Bacillus subtilis ymc encoding DNA SEQ ID NO:7.
XX
KW heterologous biological substance; Bacillus; ymc; red pigment; gene; ds.
OS Bacillus subtilis.

Key Location/Qualifiers
CDS 1..753
FT /*tag= a
FT /product= "ymc"

WO2004011609-A2.

05-FEB-2004.

25-JUL-2003; 2003WO-US023398.

26-JUL-2002; 2002US-0398853P.

(NOVO) NOVOZYMES BIOTECH INC.

Tang M, Sloma A, Sternberg D, Behr R;

WPI; 2004-143839/14.

P-PSDB; ADJ62045.

Producing a heterologous biological substance comprises cultivating pigment-deficient mutants of Bacillus cell that directs synthesis of the heterologous biological substance and has a modification of the cypX and ymc genes.

Example 1; SEQ ID NO 7; 62pp; English.

The present invention describes a method for producing a heterologous biological substance comprising cultivating a mutant of a parent Bacillus cell in a medium suitable for the production of a heterologous biological

Copyright

CC sequence represents one of the isolated P. luminescens genes

XX Sequence 705 BP; 231 A; 141 C; 118 G; 215 T; 0 U; 0 Other;

Query Match 8.2%; Score 61.6; DB 10; Length 705;
Best Local Similarity 46.4%; Pred. No. 1.6e-08;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;

```
QY 46 TTTATTCTGAGGCATTAACAGAAAACCTGACAGAAAATATTGAAACGCGCAGGCATGTT 105
DB 25 TTTACTGTCCAAAGGTGAACCTCTCGTTGTGACCAATATTCAAAAGGTGATCACGCG 84
QY 106 TTGGTGGGATCAGCCCATTAACAGCAGGTTTTCAGAGATATATTACAGATTAATT 165
DB 85 CTAATAGGATAAGCCCTTTAACTCGCGTTTTTCAAAAGACTATGTAGTGGACCTTTATT 144
QY 166 GGATGGCGGAAAGCTCAATT---TAAAGGGTTTCAGTTTACTTTCGAGGCGATGAGCG 222
DB 145 CAGTGGTCAAGTCATTAATTTCCGCAAGTCGACATATATTACCTTGTGAAGTGAAGCT 204
QY 223 GCTAATCTTCTAGAAGCGCTTGGAACTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAA 282
DB 205 TCAGCGCTTTTAGTCGCTAGTGGAAATGATAATGTTAAAGCTATCAAAAACACATCGC 264
QY 283 GAGGTATCAGAAACAGAGATTTGCAGAAAGAGCCCTTGTGGC-----TCATGGCGGG 337
DB 265 GAAATTAGAGCTCATTTACGTAACCTTTGATTTATTTCCACAGCAACATTGAAAAGT 324
QY 338 ATCGGAAGGGGATTCATAC-ATTTCTGATTTTATAGATAACAAAGCTACCAAGCTGTTG 396
DB 325 AAGCAATCAGATCATCCAAATTTAGTGACTTTTCACTAAACCATGACTACCATCTCTT 384
QY 397 AGCAAGAAGTTGAACATGCAATTTTTCGAGCAGCCTCAATTTTCGACATGCTTGTGTCAC 456
DB 385 AAAACACAAGTTGAAAACGGGTTTAATGAATCAGAACTCTTTTAAAGAAAGCTGCTTGAT 444
QY 457 ATGTCCTCGTAGAGGATTAATCGGGCGTGGCGGGCGTCAAGTTTGTATGATGAAGAAGTC 516
DB 445 ATGTCTCTTCAAGGCATTAAGGCGGACTAAAGGTAAGTCTGGGCAATCTTTTGGTCAAAAT 504
QY 517 AGTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAGCTGCGGTTTTTATC 576
DB 505 GACCTACAATTTAGTATATAAGCGTTGCCATATATTTTCGCTGAATTCCTTTTACCTC 564
QY 577 GGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCCTTGTCTTATCATCTGCTCGGTGAAG 636
DB 565 AATACCCCTCGATTACTTGGGTAAAGTATTTCAAGTTACTTTATCACCGCCCTTGCTCA 624
QY 637 CTGGGTGAGAAGTACGATCAACCATGAATTTTCTATTTGTATGCGGCGCAATCAAGGTAT 696
DB 625 ATCGAAAAGGGTTATTAAACGGTAGTTATCTCTATACAAGTAGCAGATAAACAAGTTAC 684
QY 697 CTCAATCT 704
DB 685 GGAATCGT 692
```

RESULT 3

Continuation (3 of 7) of ACF65385 from base 200001 (Photorhabdus luminescens nucleotide
WP Sequence split into 7 fragments LOCUS ACF65385 Accession ACF65385

WP	Fragment Name	Begin	End
WP	ACF65385_0	1	110000
WP	ACF65385_1	100001	210000
WP	ACF65385_2	200001	310000
WP	ACF65385_3	300001	410000
WP	ACF65385_4	400001	510000
WP	ACF65385_5	500001	610000
WP	ACF65385_6	600001	618776

Query Match 8.2%; Score 61.6; DB 10; Length 110000;
Best Local Similarity 46.4%; Pred. No. 2.1e-07;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;

```
QY 46 TTTATTCTGAGGCATTAACAGAAAACCTGACAGAAAATATTGAAACGCGCAGGCATGTT 105
DB 82443 TTTACTGTCCAAAGGTGAACCTCTCGTTGTGACCAATATTCAAAAGGTGATCACGCG 82502
QY 106 TTGGTGGGATCAGCCCATTAACAGCAGGTTTTCAGAGATATATTACAGATTAATT 165
DB 82503 CTAATAGGATAAGCCCTTTAACTCGCGTTTTTCAAAAGACTATGTAGTGGACCTTTATT 82562
QY 166 GGATGGCGGAAAGCTCAATT---TAAAGGGTTTCAGTTTACTTTCGAGGCGATGAGCG 222
DB 82563 CAGTGGTCAAGTCATTAATTTCCGCAAGTCGACATATATTACCTTGTGAAGTGAAGCT 82622
QY 223 GCTAATCTTCTAGAAGCGCTTGGAACTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAA 282
DB 82623 TCAGCGCTTTTAGTCGCTAGTGGAAATGATAATGTTAAAGCTATCAAAAACACATCGC 82682
QY 283 GAGGTATCAGAAACAGGAGATTTGCAGAAAGAGCCCTTGTGGC-----TCATGGCGGG 337
DB 82683 GAAATTAGAGCTCATTTACGTAACCTTTGATTTATTTCCACAGCAACATTGAAAAGT 82742
QY 338 ATCGGAAGGGGATTCATAC-ATTTCTGATTTTATAGATAACAAAGCTACCAAGCTGTTG 396
DB 82743 AAGCAATCAGATCATCCAAATTTAGTGACTTTTCACTAAACCATGACTACCATCTCTT 82802
QY 397 AGCAAGAAGTTGAACATGCAATTTTTCGAGCAGCCTCAATTTTCGACATGCTTGTGTCAC 456
DB 82803 AAAACACAAGTTGAAAACGGGTTTAATGAATCAGAACTCTTTTAAAGAAAGCTGCTTGAT 82862
QY 457 ATGTCCTCGTAGAGGATTAATCGGGCGTGGCGGGCGTCAAGTTTGTATGATGAAGAAGTC 516
DB 82863 ATGTCTCTTCAAGGCATTAAGGCGGACTAAAGGTAAGTCTGGGCAATCTTTTGGTCAAAAT 82922
QY 517 AGTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAGCTGCGGTTTTTATC 576
DB 82923 GACCTACAATTTAGTATATAAGCGTTGCCATATATTTTCGCTGAATTCCTTTTACCTC 82982
QY 577 GGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCCTTGTCTTATCATCTGCTCGGTGAAG 636
DB 82983 AATACCCCTCGATTACTTGGGTAAAGTATTTCAAGTTACTTTATCACCGCCCTTGCTCA 83042
QY 637 CTGGGTGAGAAGTACGATCAACCATGAATTTTCTATTTGTATGCGGCGCAATCAAGGTAT 696
DB 83043 ATCGAAAAGGGTTATTAAACGGTAGTTATCTCTATACAAGTAGCAGATAAACAAGTTAC 83102
QY 697 CTCAATCT 704
DB 83103 GGAATCGT 83110
```

RESULT 4

Continuation (32 of 57) of ACF67367 from base 3100001 (Photorhabdus luminescens nucleotide
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

WP	Fragment Name	Begin	End
WP	ACF67367_0	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000

WP	ACF67367_18	1800001	1910000	
WP	ACF67367_19	1900001	2010000	
WP	ACF67367_20	2000001	2110000	
WP	ACF67367_21	2100001	2210000	
WP	ACF67367_22	2200001	2310000	
WP	ACF67367_23	2300001	2410000	
WP	ACF67367_24	2400001	2510000	
WP	ACF67367_25	2500001	2610000	
WP	ACF67367_26	2600001	2710000	
WP	ACF67367_27	2700001	2810000	
WP	ACF67367_28	2800001	2910000	
WP	ACF67367_29	2900001	3010000	
WP	ACF67367_30	3000001	3110000	
WP	ACF67367_31	3100001	3210000	
WP	ACF67367_32	3200001	3310000	
WP	ACF67367_33	3300001	3410000	
WP	ACF67367_34	3400001	3510000	
WP	ACF67367_35	3500001	3610000	
WP	ACF67367_36	3600001	3710000	
WP	ACF67367_37	3700001	3810000	
WP	ACF67367_38	3800001	3910000	
WP	ACF67367_39	3900001	4010000	
WP	ACF67367_40	4000001	4110000	
WP	ACF67367_41	4100001	4210000	
WP	ACF67367_42	4200001	4310000	
WP	ACF67367_43	4300001	4410000	
WP	ACF67367_44	4400001	4510000	
WP	ACF67367_45	4500001	4610000	
WP	ACF67367_46	4600001	4710000	
WP	ACF67367_47	4700001	4810000	
WP	ACF67367_48	4800001	4910000	
WP	ACF67367_49	4900001	5010000	
WP	ACF67367_50	5000001	5110000	
WP	ACF67367_51	5100001	5210000	
WP	ACF67367_52	5200001	5310000	
WP	ACF67367_53	5300001	5410000	
WP	ACF67367_54	5400001	5510000	
WP	ACF67367_55	5500001	5610000	
WP	ACF67367_56	5600001	5648894	
Query Match 8.2%; Score 61.6; DB 10; Length 110000;				
Best Local Similarity 46.4%; Pred. No. 2.1e-07;				
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;				
Qy	46	TTTATTCTGAGGCATTAACAGAAAACCTGCAGAGAAATATTGAAACGGCGAGGCAATGTT	105	
Db	87577	TTTACTGTCCAAGGTGAACCTCTCGTTGTGACCAAAATATTTCAAAAGGTGATCACGCG	87636	
Qy	106	TTGGTGGGATCAGCCATTTTAACAGCAGGTTTTTCAGAGGATATATATTACAGATTAAAT	165	
Db	87637	CTAATAGGATAGCCCTTTAATCTCGGTTTTTCAAAAGACTATGTAGTGACCTTATT	87696	
Qy	166	GGATGGCGGAAAGCTCAATT---TAAAGCGTTTCAGTTTTTACTTGCAGGGCATGAGCG	222	
Db	87697	CAGTGGTCAAGTCATTATTTCCGACAAGTCACATATATTATACCTTGTGAACGTGAAGCT	87756	
Qy	223	GCTAATCTTCAGAGCGCTTGGAACCTCCGAGAGGAAAGGCTGACGAAAAGTAGAGAA	282	
Db	87757	TCAGCGCTTTTAGTCGCTAGTGGAAATGATAATGTTTAAAGCTATCAAAAACACATCGC	87816	
Qy	283	GAGTATCAGAAACAGGAGATTTGCAGAAAAGAGCCCTTTGTGGC-----TCATGGCGGG	337	
Db	87817	GAATTAGAGTCATTTACGTAACTTGATATGTTATTTCCACAGCAACATGAAAAGT	87876	
Qy	338	ATCCGAAGGCGATTTCATAC-ATTTTCTGATTTTATAGATAACAAAGCCTTACCAGCTGTTG	396	
Db	87877	AAGCAATCAGATCATCAATTTTAGTGACTTTTCACTAAACCATGACTACCAATCTCTT	87936	
Qy	397	AGACAAGAGTTGGAACATGCAATTTTTTTCAGAGCGCTCAATTTTCGACATGCTTGTGTTGAC	456	
Db	87937	AAAACACAAGTTGAAAAGCGGTTTTAATGAATCAGAAATCTTTTAAAAAAGCTGTCTTGAT	87996	
Qy	457	ATGTCTCGTGAAGCGGATAATCGGGCGTGTGGCGGGCGTTCAGTTTGTATGATGGAAGATC	516	

Db	87997	ATGTCTCTTCAAGCCATAAAAGGCGACATAAAGGTACTGGGCAATACTTTTGTCAATT	88056
Qy	517	AGTGAGGATATGCTGAATTTGGCTGGGAATATGTCTAGCTGAGCTGCCGTTTATC	576
Db	88057	GACCTACAATTAGTATATAAAGCGTTGCCATATATTTTGGCTGAAATTCCTTTTACCTC	88116
Qy	577	GGAGCTCCGATATTTTAGAGGTGGAAGAGACACTCTTCTGCTTATCATCGTCCGTGGAAG	636
Db	88117	AATACCCCTCGATTACTTGGGGTAAAGTATTTCTAGTTACTTTATCACCCCTTGTGCA	88176
Qy	637	CTGGGTGAGAAGATCAGTAACCATGAATTTTCTATTTGTATGCGCGCAATCAAGGGTAT	696
Db	88177	ATCGGAAGAGGTTTATTTAAGCGTAGTTATCTCTATACAGTACAGATAACAAAGTTAC	88236
Qy	697	CTCATTTGT 704	
Db	88237	GGAATCGT 88244	
RESULT 5			
ACF6408			
ID ACF66408 standard; DNA; 651 BP.			
XX	ACF66408;		
XX	20-NOV-2003 (first entry)		
DT	Photorhabdus luminescens nucleotide sequence #4875.		
DE	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;		
KW	detection; food; gene expression; plant; animal; microorganism; toxin;		
KW	antibiotic; biopesticide; virulence factor; disease model; plague;		
KW	whooping cough; gene; ds.		
XX	Photorhabdus luminescens.		
OS	WO200294867-A2.		
PN	28-NOV-2002.		
PD	07-FEB-2002; 2002WO-IB003040.		
XX	07-FEB-2001; 2001FR-00001659.		
PR	(INSP) INST PASTEUR.		
PA	(CNRS) CNRS CENT NAT RECH SCI.		
XX	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;		
PI	Buchrieser C;		
XX	WPI; 2003-148459/14.		
DR	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,		
XX	useful e.g. as therapeutic antimicrobials and agricultural pesticides.		
PT	Claim 28; SEQ ID NO 4875; 1205pp; French.		
PS	The invention relates to the isolation of genes and their encoded		
CC	proteins from Photorhabdus luminescens. The isolated sequences are		
CC	sources of probes and primers for detecting the genome of P. luminescens		
CC	and related species; to study polymorphisms; for gene analysis and for		
CC	detection/amplification of the genes. Antibodies (Ab) raised against the		
CC	polypeptides encoded by the genes are used for detection/identification		
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that		
CC	carry a gene-containing vector are used to select compounds that		
CC	modulate, regulate, induce or inhibit expression of the genes in plants,		
CC	animals or microorganisms other than P. luminescens and are able to alter		
CC	response or sensitivity to toxins and antibiotics produced by P.		
CC	luminescens. Cells transformed to express the genes are useful for		
CC	recombinant production of the proteins, particularly toxins and		
CC	antibacterials useful as insecticides, bactericides and fungicides. The		
CC	genes, proteins, vectors containing the genes and Ab are also useful		

Claim 27; SEQ ID NO 5263; 899pp; English.

The present invention relates to a method (M1) for identifying genes involved in plant resistance or response to pathogenic infection. M1 comprises identifying a gene whose expression is significantly altered in the incompatible interaction of plant gene expression relative to expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interfection. (M1) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to illustrate the invention.

Sequence 2000 BP; 336 A; 265 G; 284 C; 363 T; 0 U; 752 Other;

Query Match 5.4%; Score 40.4; DB 8; Length 2000;
Best Local Similarity 8.1%; Pred. No. 0.13;
Matches 29; Conservative 174; Mismatches 155; Indels 0; Gaps 0;

Qy 381 AGCCTACCAGCTGTTGAGACAAGAAGTTGAACATGCATTTTTTGAGAGCGCTCATTTTCG 440
Db WSY:::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
280 WYKMMWCTAYKKSYSRWCYMYRGCGWRGATRYWGRGYMSRMAMMYKMWYRKYGMK 339
Qy 441 ACATGCTGTGGTGGACATGCTCTGCTGAAGCGATAATCGGCCGTCGCGGGCGTCACGTTT 500
Db ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
340 RGVWAGMMMRSMCRWSKACYMYRWWRWMTTRRRRAKSSRTSRKRKRCWKMKRYKYKR 399
Qy 501 GATGATGGAAGAAGTCAGTCAGGATATGTCGAATTGCGCTGTGGAATATGTCATAGCTGA 560
Db ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
400 MRGYSRMRSCRARWMKRCBSGRAWMGCGCMTCRKMSYGMWRKSWKEMASKYWMSR 459
Qy 561 GCTGCCGTTTTTTATCCGAGCTCCGATATTTTAGAGGTGGAAGACATCTCTTGCTTA 620
Db ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
460 MYRWKXKCSRTTWGKTGCGMMGTWRCRYKKRSKMCKRCRRRWGRMYRMRWKYYMS 519
Qy 621 TCATGCTCCGTGGAAGCTGGGTGAGAAGATCAGTAACCATGAATTTTCTATTGTCATGCG 680
Db ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
520 ARTMYRCARKKYSYAARKARCWYRGKGYWAGMMMKRYKRYMYKOMMYKRYKSKCS 579
Qy 681 GC CGAATCAAGGCGTATCTCTATTGACAGAAATGGCGCAGATGCTTTCTGAGAAACGG 738
Db ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
580 WYKMSYYASCMKSARKAGAKMCKRSKMSAWSKMSRSSRCKRKCASKRSKRYNMWG 637

RESULT 7
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
FN WO200300898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
DR WPI; 2003-175290/17.

RESULT 9

AAT42063_15
Continuation (16 of 19) of AAT42063 from base 1500001 (Haemophilus influenzae complete g.
WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063

WP	Fragment Name	Begin	End
WP	AAT42063_00	1	110000
WP	AAT42063_01	100001	210000
WP	AAT42063_02	200001	310000
WP	AAT42063_03	300001	410000
WP	AAT42063_04	400001	510000
WP	AAT42063_05	500001	610000
WP	AAT42063_06	600001	710000
WP	AAT42063_07	700001	810000
WP	AAT42063_08	800001	910000
WP	AAT42063_09	900001	1010000
WP	AAT42063_10	1000001	1110000
WP	AAT42063_11	1100001	1210000
WP	AAT42063_12	1200001	1310000
WP	AAT42063_13	1300001	1410000
WP	AAT42063_14	1400001	1510000
WP	AAT42063_15	1500001	1610000
WP	AAT42063_16	1600001	1710000
WP	AAT42063_17	1700001	1810000
WP	AAT42063_18	1800001	1830121

Query Match 5.0%; Score 38; DB 2; Length 110000;

Best Local Similarity 55.2%; Pred. No. 5.6;

Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy	244	GGAACTCCGAGAGGAGGCTGACGAAAGTAAGGAAAGAGGTATCATCAGAAACAGAGA	303
Db	19210	GAATTTCAAAAATATGATTGCGAAGAACCAACGTCATAATTCGAAACGGGAAGTCGT	19269
Qy	304	TTTGCAGAAAGAGCCCTTGCGCTCATGCGGGGATCCGAAAGCGATTTCATACATTTCT	363
Db	19270	TTACCTGATACAGTCATGCTGCTGTGGCGGTGGCTCGAATGCAATTGTTGTTGCT	19329
Qy	364	GATTTTATAGATAA	377
Db	19330	GATTTTATGATGA	19343

RESULT 10

ADH10017_0
WP Sequence split into 4 fragments LOCUS ADH10017 Accession Adh10017

WP	Fragment Name	Begin	End
WP	ADH10017_0	1	110000
WP	ADH10017_1	100001	210000
WP	ADH10017_2	200001	310000
WP	ADH10017_3	300001	365186

ID ADH10017 standard; DNA; 365186 BP.

AC ADH10017;

XX XX

DT 11-MAR-2004 (first entry)

XX Human chromosome 2p21-22 fragment containing the DRIP gene.

XX human; chromosome 2; hyperplasia; tumour; thyroid; cancer; break point;
KW chromosomal band 2p21-22; DRIP; cytostatic; thymimetic;
XW chromosome 7p15; ds.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers

XX exon 1. .108

FT /*tag= a

FT /number= 1

FT 3623. .3722

FT /*tag= b

FT /number= 2

FT 3993. .4087

FT exon

FT	exon	/*tag= c
FT	exon	/number= 3
FT	exon	5085. .5215
FT	exon	/*tag= d
FT	exon	/number= 4
FT	exon	9037. .9185
FT	exon	/*tag= e
FT	exon	/number= 5
FT	exon	9305. .9337
FT	exon	/*tag= f
FT	exon	/number= 6
FT	exon	9590. .9638
FT	exon	/*tag= g
FT	exon	/number= 7
FT	exon	14110. .14297
FT	exon	/*tag= h
FT	exon	/number= 8
FT	exon	17431. .17525
FT	exon	/*tag= i
FT	exon	/number= 9
FT	exon	18796. .19016
FT	exon	/*tag= j
FT	exon	/number= 10
FT	exon	21011. .21702
FT	exon	/*tag= k
FT	exon	/number= 11
FT	exon	23046. .23224
FT	exon	/*tag= l
FT	exon	/number= 12
FT	exon	24176. .24331
FT	exon	/*tag= m
FT	exon	/number= 13
FT	exon	25528. .25650
FT	exon	/*tag= n
FT	exon	/number= 14
FT	exon	29217. .29340
FT	exon	/*tag= o
FT	exon	/number= 15
FT	exon	35653. .35804
FT	exon	/*tag= p
FT	exon	/number= 16
FT	exon	39483. .39693
FT	exon	/*tag= q
FT	exon	/number= 17
FT	exon	43699. .43834
FT	exon	/*tag= r
FT	exon	/number= 18
FT	exon	44113. .44249
FT	exon	/*tag= s
FT	exon	/number= 19
FT	exon	46670. .46828
FT	exon	/*tag= t
FT	exon	/number= 20
FT	exon	54722. .54879
FT	exon	/*tag= u
FT	exon	/number= 21
FT	exon	68050. .68159
FT	exon	/*tag= v
FT	exon	/number= 22
FT	exon	87258. .87390
FT	exon	/*tag= w
FT	exon	/number= 23
FT	exon	90303. .90416
FT	exon	/*tag= x
FT	exon	/number= 24
FT	exon	97083. .97205
FT	exon	/*tag= y
FT	exon	/number= 25
FT	exon	110713. .110804
FT	exon	/*tag= z
FT	exon	/number= 26
FT	exon	165736. .165825
FT	exon	/*tag= aa

FT exon /number= 27
FT 167807..167938
FT /*tag= ab
FT /number= 28
FT 197899..198067
FT /*tag= ac
FT /number= 29
FT 251801..251916
FT /*tag= ad
FT /number= 30
FT 275498..275592
FT /*tag= ae
FT /number= 31
FT 302825..303204
FT /*tag= af
FT /number= 32
FT 303816..303934
FT /*tag= ag
FT /number= 33
FT 304270..304342
FT /*tag= ah
FT /number= 34
FT 308977..309130
FT /*tag= ai
FT /number= 35
FT 316142..316273
FT /*tag= aj
FT /number= 36
FT 363156..363325
FT /*tag= ak
FT /number= 37
FT 364695..365186
FT /*tag= al
FT /number= 38
XX
XX WO2003093310-A1.
XX
XX 13-NOV-2003.
XX
XX 02-MAY-2003; 2003WO-EP004642.
XX
XX 01-MAY-2002; 2002DE-01019413.
XX 14-SEP-2002; 2002DE-01042705.
XX
XX (UYBR-) UNIV BREMEN.
XX
XX Bullerdiek J;
XX
XX WPI; 2003-854480/79.
XX
XX New nucleic acid from human chromosome 2, useful for treatment and
XX diagnosis of thyroid disease, especially cancer, also related
XX polypeptides and modulators.
XX
XX Claim 1; SEQ ID NO 12; 461pp; German.
XX
XX This invention describes a novel nucleic acid from human chromosome 2
XX that shows altered expression in hyperplasia and/or tumours, especially
XX of the thyroid. The invention describes a method for preparing nucleic
XX acid that can be detected in thyroid cancers where these contain an
XX aberration of the break point in chromosomal band 2p21-22. The invention
XX also describes constructs comprising a fragment of the human DRIP gene
XX and its splice variants, one of which contains all of exons 1-38 while
XX the other lacks exons 27 and 28. Some tumours contain a fusion protein of
XX DRIP, including a small segment from chromosome 3. The products of the
XX invention have cytostatic and thymimetic activity and are used for the
XX inhibition of nucleic acid expression by antisense, ribozyme or RNA
XX interference (RNAi) methods. This sequence represents the human
XX chromosome 2p21-22 DNA containing the region encoding the DRIP protein
XX encoded by exons 1-38, described in the disclosure of the invention.
XX
XX Sequence 365186 BP; 103587 A; 71408 C; 76380 G; 113811 T; 0 U; 0 Other;

Query Match 4.9%; Score 36.8; DB 10; Length 110000;
Best Local Similarity 54.4%; Pred. No. 13;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 328 CATGGCGGGATCCGAAGCGGATTTCATACATTTTCTGATTTTATAGATAACAAGCCTAC 387
DB 107307 CATAGAGTTTACTCTGTGACCAAGCAATACTACTTCTAGGTATATAGTCATGAGAAATGAA 107366
QY 388 CAGCTGTTGAGACAGAAGTTGAACATGCTATTTTGGACAGCCTCATTTTCGACATGCT 447
DB 107367 AACATGTTTACACAAAACCTTGATCATGAATATTTCTAGCAAAATATTCTATAATATTC 107426
QY 448 TGTTCGACATGTCTC 463
DB 107427 AAATTGGAATAGCCC 107442
RESULT 11
ADH10017.1
Continuation (2 of 4) of ADH10017 from base 100001 (Human chromosome 2p21-22 fragment con
WP Sequence split into 4 fragments LOCUS ADH10017 Accession Adh10017
WP Fragment Name Begin End
WP ADH10017_0 1 110000
WP ADH10017_1 100001 210000
WP ADH10017_2 200001 310000
WP ADH10017_3 300001 365186
Query Match 4.9%; Score 36.8; DB 10; Length 110000;
Best Local Similarity 54.4%; Pred. No. 13;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 328 CATGGCGGGATCCGAAGCGGATTTCATACATTTTCTGATTTTATAGATAACAAGCCTAC 387
DB 7307 CATAGAGTTTACTCTGTGACCAAGCAATACTACTTCTAGGTATATAGTCATGAGAAATGAA 7366
QY 388 CAGCTGTTGAGACAGAAGTTGAACATGCTATTTTGGACAGCCTCATTTTCGACATGCT 447
DB 7367 AACATGTTTACACAAAACCTTGATCATGAATATTTCTAGCAAAATATTCTATAATATTC 7426
QY 448 TGTTCGACATGTCTC 463
DB 7427 AAATTGGAATAGCCC 7442
RESULT 12
AAQ75343
ID AAQ75343 standard; cDNA to mRNA; 3048 BP.
XX
XX AAQ75343;
XX
XX 16-OCT-1995 (first entry)
XX
XX DT
XX DE
XX DE Soybean phosphoenolpyruvate carboxylase cDNA to mRNA.
XX KW Soybean; phosphoenolpyruvate carboxylase; protein production;
XX KW fat production; transgenic crop plants; ds.
XX OS Glycine max.
XX
XX Key Location/Qualifiers
XX CDS 28..2931
XX /*tag= a
XX
XX JF06319567-A.
XX
XX PD 22-NOV-1994.
XX
XX PF 27-SEP-1991; 91JP-00274950.
XX
XX PR 27-SEP-1991; 91JP-00274950.
XX
XX (NORQ) NORINSUISANSHO SHOKUHIN SOGO.
XX PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
XX PA

XX WPI; 1995-040320/06.
DR P-PSDB; AAR67428.
XX Phospho-enol pyruvate carboxylase gene derived soy beans - to provide
PT increased protein or fat production in transgenic crop plants.
XX Claim 1; Page 5-11; 19pp; Japanese.
XX
CC AAQ75343 encodes AAR67428 soybean phosphoenolpyruvate carboxylase (PEPC).
CC Expression of PEPC increases protein prodn., whilst inhibition of PEPC
CC expression increases fat prodn. Therefore PEPC can be used to modulate
CC protein and fat prodn. in transgenic crop plants
XX
SQ Sequence 3048 BP; 876 A; 605 C; 724 G; 843 T; 0 U; 0 Other;
Query Match 4.8%; Score 36.4; DB 2; Length 3048;
Best Local Similarity 51.9%; Pred. No. 2.9;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 502 ATGATGGAAGAAGTCAGTCAGGATATGCTGGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2446 ATGCTGCAAGAGATGTACATCAATGCGCTTCTTTAGGCTCACACTTGATTTGGTGAA 2505
QY 562 CTGCGGTTTTTATCGGAGCTCGGATATTTAGAGTGTGGAAGACACACTCTTGTCTTAT 621
Db 2506 ATGGTGTGTTGCCAAGAGATCGAAATTTGCGCTCTGTAATGATAGACTCTTGTTTCA 2565
QY 622 CATCGTCCGTGGAGCTGGGTGAGAAGATCAGTAACCA 659
Db 2566 AAGGATCTGTGCGGTTGGGATCAATTGAGGAACAA 2603
RESULT 13
AAQ75344
ID AAQ75344 standard; cDNA to mRNA; 3211 BP.
AC AAQ75344;
XX
XX 16-OCT-1995 (first entry)
DT
XX
DE Soybean phosphoenolpyruvate carboxylase cDNA to mRNA.
XX
XX Soybean; phosphoenolpyruvate carboxylase; protein production;
KW fat production; transgenic crop plants; ds.
XX
XX Glycine max.
XX
XX Key Location/Qualifiers
FH 117..3020
FT /*tag= a
FT
XX
XX JP06319567-A.
XX
XX 22-NOV-1994.
XX
XX 27-SEP-1991; 91JP-00274950.
XX
XX 27-SEP-1991; 91JP-00274950.
XX
XX (NORQ) NORINSUISANSHO SHOKUJIN SOGO.
PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
XX
XX WPI; 1995-040320/06.
DR P-PSDB; AAR67429.
XX
XX Phospho-enol pyruvate carboxylase gene derived soy beans - to provide
PT increased protein or fat production in transgenic crop plants.
XX
XX Claim 1; Page 13-19; 19pp; Japanese.
XX
CC AAQ75344 encodes AAR67429 soybean phosphoenolpyruvate carboxylase (PEPC).
CC Expression of PEPC increases protein prodn., whilst inhibition of PEPC

CC expression increases fat prodn. Therefore PEPC can be used to modulate
CC protein and fat prodn. in transgenic crop plants
XX
SQ Sequence 3211 BP; 910 A; 651 C; 747 G; 903 T; 0 U; 0 Other;
Query Match 4.8%; Score 36.4; DB 2; Length 3211;
Best Local Similarity 51.9%; Pred. No. 3;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 502 ATGATGGAAGAAGTCAGTCAGGATATGCTGGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2535 ATGCTGCAAGAGATGTACATCAATGCGCTTCTTTAGGCTCACACTTGATTTGGTGAA 2594
QY 562 CTGCGGTTTTTATCGGAGCTCGGATATTTAGAGTGTGGAAGACACACTCTTGTCTTAT 621
Db 2595 ATGGTGTGTTGCCAAGAGATCGAAATTTGCGCTCTGTAATGATAGACTCTTGTTTCA 2654
QY 622 CATCGTCCGTGGAGCTGGGTGAGAAGATCAGTAACCA 659
Db 2655 AAGGATCTGTGCGGTTGGGATCAATTGAGGAACAA 2692
RESULT 14
ABA03041_24/c
Continuation (25 of 30) of ABA03041 from base 2400001 (Listeria monocytogenes EGD-e genom
WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041
WP Fragment Name Begin End
WP ABA03041_00 1 110000
WP ABA03041_01 100001 210000
WP ABA03041_02 200001 310000
WP ABA03041_03 300001 410000
WP ABA03041_04 400001 510000
WP ABA03041_05 500001 610000
WP ABA03041_06 600001 710000
WP ABA03041_07 700001 810000
WP ABA03041_08 800001 910000
WP ABA03041_09 900001 1010000
WP ABA03041_10 1000001 1110000
WP ABA03041_11 1100001 1210000
WP ABA03041_12 1200001 1310000
WP ABA03041_13 1300001 1410000
WP ABA03041_14 1400001 1510000
WP ABA03041_15 1500001 1610000
WP ABA03041_16 1600001 1710000
WP ABA03041_17 1700001 1810000
WP ABA03041_18 1800001 1910000
WP ABA03041_19 1900001 2010000
WP ABA03041_20 200001 2110000
WP ABA03041_21 2100001 2210000
WP ABA03041_22 2200001 2310000
WP ABA03041_23 2300001 2410000
WP ABA03041_24 2400001 2510000
WP ABA03041_25 2500001 2610000
WP ABA03041_26 2600001 2710000
WP ABA03041_27 2700001 2810000
WP ABA03041_28 2800001 2910000
WP ABA03041_29 2900001 2944528
Query Match 4.8%; Score 36.2; DB 6; Length 110000;
Best Local Similarity 45.8%; Pred. No. 21;
Matches 128; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY 101 ATGTTTTGGTGGGATCAGCCCATTTTAAACAGCAGGTTTTTACAGAGGATTTATATTACAGAT 160
Db 25556 AGGATTTGGAGGCGGGGATTAATGGAGAAAGGTTTACTACGGAATTAATGTTTACAT 25497
QY 161 TAATTGGATGGCGAAAGCTCAATTTAAAGCGTTTCAGTTTTTACTTTGCGAGGCGATGAGG 220
Db 25496 TAACTAATAATCCAAAGTTTCAAGAAGTAGCAGAAATGGTATATTGGAACGAGGAACAA 25437
QY 221 CGGCTAATCTTCTAGAGCGCTTGGAACTCCGAGAGGAAAGGCTGACGAAAGATGAGGA 280
Db 25436 TTGTTCAATTGTCAGAAACAAATGAGTATGTGATGCAATTGACGAGGAAGTGTAAACGC 25377

QY 281 AAGAGGTATCAGCAACAGGAGATTTCAGAAAGAGCCCTTGTGGCTCATGGCGGGATC 340
|||
Db 25376 AATGTCTGCAAGAAACAGTTCTTGGGAAAGGAAGCATCTCGGTTTACTCGAGAAG 25317
|||
QY 341 CGAAGCGGATTTCATACATTTCTCGATTTTATAGATAACAAA 381
|||
Db 25316 AGACAAAATCCAAACCCCTTCTAAACGGTAGCCTGGAAA 25276
|||

RESULT 15
ACF72025
ID ACF72025 standard; DNA: 1338 BP.
XX
AC ACF72025;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #10492.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
detection; food; gene expression; plant; animal; microorganism; toxin;
antibiotic; biopesticide; virulence factor; disease model; plague;
whooping cough; gene; ds.
XX
OS Photorhabdus luminescens.
XX
FN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-15003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 10492; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens genes
XX

SQ Sequence 1338 BP; 344 A; 269 C; 367 G; 358 T; 0 U; 0 Other;
Query Match 4.8%; Score 36; DB 10; Length 1338;

Best Local Similarity 51.9%; Pred. No. 2.6;
Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 392 TGTGAGACAAGAAGTTGAACATGCATTTTTCAGCAGCCTCATTTTCGACATGCTTGT 451
|||
Db 590 TGC CGGAACAACCTCCTTGAGTCAGAACTGTTTGGTCATGCTAAAGGTGCATTTACCGGTG 649
|||
QY 452 TGSACATGTCGTGAAGCGATAATCGGGCGTSCGCGGGCGTCAGTTTTCATGATGGAAG 511
|||
Db 650 CGGTCAGCAGCCGTGAAGGTTATTCTTGTGCGCAGGGGGGAACGCTGTTTCTTGATG 709
|||
QY 512 AAGTCAGTCAGGATATGCTGAATTTTGGCTGTGGAAT 547
|||
Db 710 AAATCGTGTATATGCCACAGGCTTTTCAGGTTAAAT 745
|||

Search completed: December 10, 2004, 13:46:26
Job time : 424.058 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:19:06 ; Search time 76.1418 Seconds
(without alignments)
7029.309 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagatgagatgccggaat.....aacggatcacatctgaagga 753.

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 35539441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48.6	6.5	7218	1	US-08-232-463-14
2	41.2	5.5	1659	1	Sequence 14, Appli
3	41.2	5.5	1659	1	Sequence 1, Appli
4	38	5.0	1830121	4	US-08-231-729B-2
5	38	5.0	1830121	4	Sequence 1, Appli
6	38	5.0	1830121	4	Sequence 1, Appli
7	36.2	4.8	399	4	US-09-557-884-1
8	34.8	4.6	640681	4	US-09-643-990A-1
9	34	4.5	395	4	US-10-329-960-1
10	33.8	4.5	1230025	4	US-09-621-976-8976
11	33.8	4.5	1664976	4	Sequence 1, Appli
12	33.8	4.5	1664976	4	Sequence 45, Appli
13	33.4	4.4	879	4	US-09-198-452A-1
14	33.2	4.4	337	4	US-08-916-421B-1
15	33.2	4.4	337	4	US-09-692-570-1
16	32.6	4.3	1500	4	US-08-489-039A-3818
17	32.4	4.3	234	4	US-03-270-767-2403
18	32.4	4.3	1467	4	Sequence 17685, A
19	32.4	4.3	1485	3	Sequence 545, App
20	32.4	4.3	2291	4	Sequence 6410, App
21	32.4	4.3	2980	4	Sequence 2187, Ap
22	32.4	4.3	2980	4	Sequence 627, App
23	32	4.2	1141	4	Sequence 2187, Ap
24	31.8	4.2	1882	1	Sequence 4242, Ap
25	31.8	4.2	1882	1	Sequence 4013, Ap
26	31.8	4.2	1882	1	Sequence 22, Appli
27	31.8	4.2	1882	2	Sequence 15, Appli
					Sequence 29, Appli

Query Match 6.5%; Score 48.6; DB 1; Length 7218;

Sequence 15, Appli
Sequence 29, Appli
Sequence 16, Appli
Sequence 6, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 78, Appli
Sequence 2813, Ap
Sequence 1, Appli
Sequence 4062, Ap
Sequence 4041, Ap
Sequence 18033, A
Sequence 809, App
Sequence 2129, Ap
Sequence 842, App
Sequence 22, Appli
Sequence 2971, Ap
Sequence 3567, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgt-File
US-08-232-463-14

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1659 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1659
OTHER INFORMATION: for codons 28, 32, 112, 130, 142, 190, 212,
OTHER INFORMATION: 217, 222, 266, 329, 336, 396, 436, 512, and 532, if the 3'
OTHER INFORMATION: nucleotide is T or C, then the 5' nucleotide is C; and if the
OTHER INFORMATION: nucleotide is A, then the 3' nucleotide is A or G
US-08-231-729B-2

Query Match 5.5%; Score 41.2; DB 1; Length 1659;
Best Local Similarity 30.08; Pred. No. 0.0062;
Matches 93; Conservative 39; Mismatches 178; Indels 0; Gaps 0;

QY 304 TTTCAGAAAGAGCCCTTGTGGCTCATGCGGGGATCCGAAGCGGATTTCATACATTTTCT 363
DB 1255 TCNGNGAYATHGGNTAYTTVGAYGARGAYGNCAYGTNTAYATHGTNGAYMGNYTNAAR 1314
QY 364 GATTTTATAGATACAAAGCCCTACAGCTGTTGAGACAAGAGTTGAACATGCAATTTT 423
DB 1315 TCNYTNAATHAARTAYAAAGGNTAYCARGTNCNCNGCNGARYTNGARGCNYTNYTNTN 1374
QY 424 GAGCAGCCTCATTTTCAGCATGCTTGTGACATGCTCGTGAAGCGATAATCGGGCGT 483
DB 1375 CARCAYCCTTATGARGAYGCGNGTNGCNGGNTCNCNGAYGARGTNGCNGNGAY 1434
QY 484 GCGGGGCGCTCAGTTTGATGATGGAAGAAGTCAGTGAGGATATGCTGAATTTGGCTGTG 543
DB 1435 YTNCCNGGCGNGTNGTNGTNTAARGARGAARTCNATHACNGARAARGARATHCAR 1494
QY 544 GAATATGCTAGCTAGCTGCGCTTTTATCGGAGCTCGGATATTTAGAGGTGAA 603
DB 1495 GAYTAYTNGGNGCARGTACNACNTCNTNAARAARYTNMGNGGNGTNGARTTYGTN 1554
QY 604 GAGACATCC 613
DB 1555 AARGATNC 1564

RESULT 4
US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 5.0%; Score 38; DB 4; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 2.9;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 244 GGAACCTCGAGAGAGGCTGAACGAAAGTAAGGAAAGAGGTATCACGAAACAGGAGA 303
DB 1519210 GAATTTCAAAATGATTGGCGAAGAACCAACGTCAAATTCGAACGGAAGGTCT 1519269
QY 304 TTTCAGAAAGAGCCCTTGTGGCTCATGCGGGGATCCGAAGCGGATTCATACATTTTCT 363
DB 1519270 TTACCTGATACAGTCATTGCTGCTGTTGGCGGTGCTCGAATGCAATGCTATGTTGCT 1519329
QY 364 GATTTATAGATAA 377
DB 1519330 GATTTTATGATGA 1519343

RESULT 5
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match          5.0%; Score 38; DB 4; Length 1830121;
Best Local Similarity 55.2%; Pred.No. 2.9;
.Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 244 GGAACCCGAGAGGAAAGGCTGAACGAAAGTAAGGAAAGAGGTATCATCAGAAACAGGAGA 303
Db 1519210 GAATTTCAAAATATGATTGGCGAAGAACCAACGTCARATTTCTCGAACGGGAAGTCTGT 1519269

QY 304 TTTCGAGAAAGCCCTTTGGCTCATGGCGGGATCCGAAGCGGATTCTATACATTTTCT 363
Db 1519270 TTACCTGATACAGTCATGCTGCTGTTGGCGGTGCTCGAATGCAATTGGTATGTTTGTCT 1519329

QY 364 GATTTTATAGATTA 377
Db 1519330 GATTTTATTGATGA 1519343

RESULT 6
US-10-329-960-1
; Sequence 1, Application US/10329960
; Patent No. 6742927
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; Patent No. 6742927
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PBI86P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

```
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature

Query Match      5.0%; Score 38; DB 4; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 2.9;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 244 GGAACCTCCGAGAGGAGCTGACGAAAGTAGGAAAGAGGTATCAGCAACAGAGA 303
Db 1519210 GAATTTCAAAAAATGATTGGCGAAGAAACCAACGTCAAATTCGAAACGGGAAGGTCGT 1519269

Qy 304 TTTGCAGAAAGAGCCCTTGTGGCTCATGGCGGGGATCCGAAGGCGATTACATATTTC 363
Db 1519270 TTACCTGATACAGTCATTTGCTGCTGTTGGCGGTGCGATGCAATGTGTAATTGCT 1519329

Qy 364 GATTTTATAGATAA 377
Db 1519330 GATTTTATTGATGA 1519343

RESULT 7
US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-8976

Query Match      4.8%; Score 36.2; DB 4; Length 399;
Best Local Similarity 10.0%; Pred. No. 0.12;
Matches 35; Conservative 158; Mismatches 156; Indels 0; Gaps 0;

Qy 357 ATTTTCTGATTTTATAGATAACAAAGCCTACAGCTGTGAGACAAGATTGAACATGC 416
Db 356 AWCKWWSGAMKWRGTGKWKYVYMTSRGSRYRYKTSAMMGRAMKRRKTKKMYMKGG 297

Qy 417 ATTTTGTGACAGCCTCATTTTCGACATGCTGTTTGGACATGCTCGTGAAGCGATAAT 476
Db 296 GKXGSGTYMAMRSRRSGTGRWSYRAMWRGSKWGGSYRMAGYRSSRWRSWYSAMWRK 237

Qy 477 CGGGCGTGGCGGGCGCTCAGTTTGATGAGGAAGTCAAGTGAGGATATGCTGAATTT 536
Db 236 KMTCKWGRSSWGRSTGYAMMYKSWCTSRKMYKKRKKWRKCTCTKTCYRGST 177

Qy 537 GGCTGTGGAATATGTCATAGCTGAGCTGCCGTTTATTCGAGCTCCGATATTTTGA 596
Db 176 YKWKAYYTKRKRKWRWTWYVYKSYMSMKKTWRMKTAYVYWKWKMTKTKWTCTMCWK 117

Qy 597 GGTGGAAGACACTCCTTCCTTATCATCGTCGTGGAACCTGGTGAGAGATCAGTAA 656
Db 116 CTTYMAGTMMYRYRYRYVYAKRAKWSKRCTWTSTTTCYMKYNAKCKWSYWSMSMMKWK 57
```



```
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

Query Match          4.5%; Score 33.8; DB 4; Length 1230025;
Best Local Similarity 47.1%; Pred. No. 51;
Matches 104; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 401 AGAAGTTGACATGCTATTTTGACGAGCCTCATTTTCGACATGCTTTGCGACATGT 460
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
725555 AAGAAGCTAAATCAAGATTAGAAAAGTTACACAGAAAATACCTGTTGCAAAATC 725496

QY 461 CTCGTGAAGCATAAATCGGGCGTGCAGCGGGCGCTCAGTTTGATCGAAGAGTCAGTG 520
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
725495 GCCTTCTGAGAAATTACACAAATTCGACACCAAGTGATATATCGACGAGATAAAA 725436

QY 521 AGGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAGCTGCCGTTTTTTATCGGAG 580
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
725435 AAGAAGCTGCTTCAAGGTGACAGACAGAAAATTAGTGAAGGCGCTGTTATGCTATG 725376

QY 581 CTCGGATATTTAGAGTGGAGAGACACTCCTTGCTTAT 621
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
725375 AGCATAAATTAACAGCTGGAAGAACACTACAGCGTTAT 725335

RESULT 11
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; FILE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
```

```
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664855)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
```

Query Match 4.5%; Score 33.8; DB 4; Length 1664976;
Best Local Similarity 49.2%; Pred. No. 58;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
QY 498 TTGTGATGGAAGAAGTCAGTCAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGC 557
|||
Db 734773 TTGGTTTATTTGCTGATGAAGTCATAGTATGCTGTATGTTGTAGGTATAGGAACGGTAA 734832
|||
QY 558 TGAGCTGCCGTTTTTTTATCGGAGCTCCGGATATTTTAGAGGTTGGAAGACACTCCCTTGC 617
|||
Db 734833 TGAGAGGCATTTTACAAAGAGGCTGAGGAGATTTTAAATAAAGTGGATTTAATAGTGTG 734892
|||
QY 618 TTATCATCGTCCGTGGAAGCTGGGTGAGAAGATCAGTAACCATGAATTTTCTATTGTAT 677
|||
Db 734893 TTATAAAATTACAAAAGTTTGTTCGAGAGGCTTAACAAGCCCAATATATACAACCTGGAAT 734952
|||
QY 678 G 678
Db 734953 G 734953
```

RESULT 12

```
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
```

```

; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10398)..(10398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature

```

```

; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (741694)..(741694)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)

```

; OTHER INFORMATION: n equals a, t, c, or g

Query Match 4.4%; Score 33.8; DB 4; Length 1664976;
Best Local Similarity 49.2%; Pred. No. 58;
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 498 TTTGATGATCGAAGAGTCACTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGC 557
Db 734773 TTGGTTATTGTGTGATGAAGTATAGTATGCTGTATGTTGTAGGTATAGGAAGCGGTAA 734832

QY 558 TGAGCTGCCGCTTTTATCCGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCCCTTGC 617
Db 734833 TGAGAGCAATTTACAAAAGAGCTGAGAGATTTTAAATAAGTGAATTAATAGTGTG 734892

QY 618 TTATCATGCTCGTGGAGCTGGGTGAGAAGATCAGTAAACCATGAATTTTCTATTGTAT 677
Db 734893 TTATAAAATTTACAAAAGTTTGTGAGAGGCTTAAACAGCCAATATATACAACCTGGAAT 734952

QY 678 G 678
Db 734953 G 734953

RESULT 13

US-09-489-039A-3818
; Sequence 3818, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3818
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-3818

Query Match 4.4%; Score 33.4; DB 4; Length 879;
Best Local Similarity 49.2%; Pred. No. 1.4;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 473 TAATCGGCGTGC CGCGCGCTCAGTTTGATGATGGAAGATCAGTCAGTCAGGATATGCTGA 532
Db 695 TCAGCGCGCGACGCGCGGAATGGGGGTGATGATCAACGACGCCCGCAGTATATCTGGA 754

QY 533 ATTTGGCTGTGGAATATGTCATAGCTCAGCTGCCGTTTTTTATCGAGCTCCGATATTT 592
Db 755 CCAGCGCGTGC AAATGTTCTGCCGGGCTGGCGCTGTTCAATCAGCGTATGCTTTTA 814

QY 593 TAGAGGTGGAAGAGACACTCTTGTTCATCATCGTCGCGGAAGCTGGGTGAGAAGATC 651
Db 815 ACCTGCTGGCGAGCGCGCTGCGGATCGTCTCGATCCCGACCTGATCGCGGAGCATAGC 873

RESULT 14

US-09-270-767-2403/c
; Sequence 2403, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2403

; LENGTH: 337
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-2403

Query Match 4.4%; Score 33.2; DB 4; Length 337;
Best Local Similarity 54.0%; Pred. No. 0.98; 58; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 93 GCAGCGCATGTTTGGTGGGATCAGCCCATTTAAACAGCAGGTTTTTCAGAGGATTATAT 152
Db 260 GCCTCGGGTGTGTTTGAAGGTATCAACGCATTTTACATATTTATATACATCATTTGTT 201

QY 153 TTACAGATTAAATTGGATGGCGGAAGCTCAATTTAAAGCGTTTTCAGTTTACTTGCAGG 212
Db 200 CAAGAGATAGAGATCGAGAGGCGGCGAGTTTGGAAAAGATGTTACTTGAAGGTGCATA 141

QY 213 GCATGA 218
Db 140 GGATAA 135

RESULT 15

US-09-270-767-17685/c
; Sequence 17685, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17685
; LENGTH: 337
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-17685

Query Match 4.4%; Score 33.2; DB 4; Length 337;
Best Local Similarity 54.0%; Pred. No. 0.98; 58; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 93 GCAGCGCATGTTTGGTGGGATCAGCCCATTTAAACAGCAGGTTTTTCAGAGGATTATAT 152
Db 260 GCCTCGGGTGTGTTTGAAGGTATCAACGCATTTTACATATTTATATACATCATTTGTT 201

QY 153 TTACAGATTAAATTGGATGGCGGAAGCTCAATTTAAAGCGTTTTCAGTTTACTTGCAGG 212
Db 200 CAAGAGATAGAGATCGAGAGGCGGCGAGTTTGGAAAAGATGTTACTTGAAGGTGCATA 141

QY 213 GCATGA 218
Db 140 GGATAA 135

Search completed: December 10, 2004, 18:17:34
Job time : 98.1418 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 16:12:37 ; Search time 447.285 Seconds
(without alignments)
9281.958 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 GTGAATGAGATGACCGGAAT.....AACGGATCACATCTGAAGGA 753

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	753	100.0	753	16	US-10-627-124-7
2	38	5.0	1194	16	US-10-282-122A-22329
3	38	5.0	1830121	14	US-10-329-960-1
4	38	5.0	1830121	16	US-10-329-960-1
5	38	5.0	1830121	18	US-10-158-875-1
6	37.4	5.0	2746	18	US-10-425-115-59495
7	36.4	4.8	945	16	US-10-425-114-29941
8	36.4	4.8	3179	16	US-10-425-114-29880
9	36.4	4.8	3285	16	US-10-425-114-11232
10	36.4	4.8	3542	16	US-10-424-599-96841
11	35.4	4.7	736	16	US-10-296-115-256
12	35.4	4.7	3681	16	US-10-424-599-6970

13	35.4	4.7	4061	10	US-09-814-353-21372	Sequence 21372, A
14	35.4	4.7	13555	10	US-09-764-891-7281	Sequence 7281, A
15	35.4	4.7	1163020	16	US-10-398-221-10	Sequence 10, Appl
16	35.4	4.7	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
17	35.2	4.7	1056	18	US-10-653-047-5346	Sequence 5346, Ap
18	35.2	4.7	3117	17	US-10-437-963-10898	Sequence 10898, A
19	35	4.6	65787	16	US-10-052-482-109	Sequence 109, App
20	35	4.6	105499	17	US-10-322-281-326	Sequence 326, App
21	34.8	4.6	1355	16	US-10-425-114-32415	Sequence 32415, A
22	34.8	4.6	2509	16	US-10-425-114-24957	Sequence 24957, A
23	34.8	4.6	2791	18	US-10-425-115-63082	Sequence 63082, A
24	34.8	4.6	640681	9	US-09-790-988-1	Sequence 1, Appli
25	34.6	4.6	513	9	US-09-867-701-2391	Sequence 2391, Ap
26	34.6	4.6	526	16	US-10-424-599-92180	Sequence 92180, A
27	34.6	4.6	1984	16	US-10-881-088-33	Sequence 33, Appl
28	34.6	4.6	3931	15	US-10-006-780-1	Sequence 1, Appli
29	34.2	4.5	250	11	US-09-987-899-5354	Sequence 5354, Ap
30	34.2	4.5	1446	16	US-10-282-122A-18304	Sequence 18304, A
31	34	4.5	395	9	US-09-894-844-45	Sequence 45, Appl
32	34	4.5	395	16	US-10-388-902-45	Sequence 45, Appl
33	34	4.5	395	16	US-10-647-089-45	Sequence 45, Appl
34	34	4.5	2325	16	US-10-424-599-91791	Sequence 91791, A
35	33.8	4.5	779	13	US-10-027-632-3731	Sequence 3731, Ap
36	33.8	4.5	779	15	US-10-027-632-3731	Sequence 3731, Ap
37	33.8	4.5	875	10	US-09-814-353-20346	Sequence 20346, A
38	33.8	4.5	963	14	US-10-022-832-27	Sequence 27, Appl
39	33.8	4.5	963	16	US-10-282-122A-18554	Sequence 18554, A
40	33.8	4.5	1442	13	US-10-098-841-145	Sequence 145, Appl
41	33.8	4.5	1698	17	US-10-250-615-34	Sequence 34, Appl
42	33.8	4.5	4128	18	US-10-425-115-94844	Sequence 94844, A
43	33.8	4.5	1230025	16	US-10-289-782-1	Sequence 1, Appli
44	33.6	4.5	666	13	US-10-027-632-261431	Sequence 261431, A
45	33.6	4.5	666	15	US-10-027-632-261431	Sequence 261431, A

ALIGNMENTS

RESULT 1

US-10-627-124-7
; Sequence 7, Application US/10627124
; Publication No. US20040096944A1
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient Mutants Of Bacillus Cells
; FILE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-627-124-7

Query Match	100.0%;	Score 753;	DB 16;	Length 753;
Best Local Similarity	100.0%;	Pred. No. 1.4e-226;		
Matches 753;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GTGAATGAGATGACCGGAATGTTAAACGGAAGAGGCTGTGCATTTTATTTCGTGAGCA	60	
Db	1	GTGAATGAGATGACCGGAATGTTAAACGGAAGAGGCTGTGCATTTTATTTCGTGAGCA	60	
Qy	61	TTACAGAAAACCTGCAGAGAAATATTTTAAACGGCGCAGGATGTTTGGTGGGATCAGC	120	
Db	61	TTACAGAAAACCTGCAGAGAAATATTTTAAACGGCGCAGGATGTTTGGTGGGATCAGC	120	
Qy	121	CCATTTTAAACGAGGTTTTCAGAGGATATATTTTACAGATTAATTCAGATGGCGGAAAGCT	180	
Db	121	CCATTTTAAACGAGGTTTTCAGAGGATATATTTTACAGATTAATTCAGATGGCGGAAAGCT	180	

appi.out


```
;/ NAME/KEY: misc feature
;/ LOCATION: (29298)..(29298)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (36543)..(36543)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (36551)..(36551)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (36636)..(36636)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (40808)..(40810)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (44416)..(44416)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (44905)..(44905)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (44975)..(44975)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (45593)..(45593)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (45732)..(45732)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (47036)..(47036)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51334)..(51334)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51602)..(51602)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51786)..(51786)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51805)..(51805)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (55369)..(55369)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (65309)..(65309)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (65313)..(65313)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (80024)..(80024)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (100091)..(100091)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (102696)..(102696)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (105121)..(105121)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (107248)..(107248)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (117136)..(117136)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (119750)..(119750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (119924)..(119924)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (120038)..(120038)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (121344)..(121344)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (122167)..(122167)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (122336)..(122336)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (131340)..(131340)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (131360)..(131360)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (139910)..(139910)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (140398)..(140398)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (142750)..(142750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (145058)..(145058)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (145171)..(145171)
```

```

; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152530)..(152530)

```

Query Match 5.0%; Score 38; DB 14; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 57;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy	244	GGNACTCCGAGAGCAAGCTGAAACGAAAAGTAAGGAAAGAGGTATCACGGAACAGGAGA	303
Dβ	1519210	GAATTTCAAAAAATGATTGGCGAAGAAACCAAACTGCAAACTCTCGAAACGGGAAGGTCTGT	1519269

Qy	304	TTTGCAGAAAGAGCCCTTGTGGCTCATGCGCGGGATCCGAAGCGCATTCATACATTTTCT	363
Dp	1519270	TTTACCTGATACAGTCAATGCTGCTGTGGCGGTGGCTCGAAATGCAATGGTATGTTTGCT	1519329

Qy 364 GATTTATAGATAA 377
|||
Db 1519330 GATTTATTGATGA 1519343

```

RESULT 4
US-10-329-670-1
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PS186P1
; CURRENT APPLICATION NUMBER: US/10/329,670
; CURRENT FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c

```

```
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122167)..(122167)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (122336)..(122336)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131340)..(131340)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (131360)..(131360)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (139910)..(139910)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (140398)..(140398)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (142750)..(142750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145058)..(145058)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (152530)..(152530)

Query Match      5.0%; Score 38; DB 16; Length 1830121;
Best Local Similarity 55.2%; Pred.No. 57;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy      244  GGAACCTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAAGAGGTATCACGAAACAGGAGA 303
Db      1519210  GAATTTCAAAAATGATTGGCGAAGAAACCAACAGTCAAAATTTCTCGAACCGGAAGGTCGT 1519269

Qy      304  TTTCGAGAAAGAGCCCTTGTGGCTCATGGCGGGATCCGAGGGGATTACATATTTTCT 363
Db      1519270  TTACCTGATACAGTCATTGCTGCTGTGGCGGTGGCTCGAATTCGAATTTGTTATGTTTGTCT 1519329

Qy      364  GATTTTATAGATAA 377
Db      1519330  GATTTTATGATGA 1519343

RESULT 5
US-10-158-865-1
; Sequence 1, Application US/10158865
; Publication No. US2004020309A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P2C1D1
; CURRENT APPLICATION NUMBER: US/10/158,865
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
```

```
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (29298)..(29298)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (36543)..(36543)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (36551)..(36551)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (36636)..(36636)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (40808)..(40810)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (44416)..(44416)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (44905)..(44905)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (44975)..(44975)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (45593)..(45593)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (45732)..(45732)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (47036)..(47036)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51334)..(51334)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51602)..(51602)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51786)..(51786)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (51805)..(51805)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (55369)..(55369)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (65309)..(65309)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (65313)..(65313)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (80024)..(80024)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (100091)..(100091)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (102696)..(102696)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (105121)..(105121)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (107248)..(107248)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (117136)..(117136)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (119750)..(119750)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (119924)..(119924)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (120038)..(120038)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (121344)..(121344)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (122167)..(122167)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (122336)..(122336)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (131340)..(131340)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (131360)..(131360)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (139910)..(139910)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (140398)..(140398)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (142750)..(142750)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (145058)..(145058)
;/ OTHER INFORMATION: n equals a,t,c, or g
;/ FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

Query Match      5.0%; Score 38; DB 18; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 57;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 244 GGAATCCGAGAGAAAGGCTGACGAAAGTAGGAAAGAGGATATCAGCAAAACAGAGA 303
Db 1519210 GAATTCAAAATAATGATGGCGAGAAACCAAAAGTCAAAATCTCGAAGCGGAAGTCT 1519269

Qy 304 TTGCGAGAAAGAGCCCTTGCTGCTCATGGCGGGGATCCGAGGCGATTCATACATTTTCT 363
Db 1519270 TTACCTGATACAGTCATGCTGCTGTTGGCGGTGGCTCGAATGCAATGCTATGTTGCT 1519329

Qy 364 GATTTATAGATAA 377
Db 1519330 GATTTATAGATGA 1519343

RESULT 6
US-10-425-115-59495/c
; Sequence 59495, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 59495
; LENGTH: 2746
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154258C.1
US-10-425-115-59495

Query Match      5.0%; Score 37.4; DB 18; Length 2746;
Best Local Similarity 45.9%; Pred. No. 2.2;
Matches 128; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 457 ATGCTCGTGAAGCGATAATCGGCGGTGCGGGCGGTGCTGCTGATGATGGAAGATC 516
Db 2404 ATGTCATTGGATTAAACCAATGCACCAAGCTATTTTCATGATTTGATGAATAGTGT 2345

Qy 517 AGTGAGGATATGCTGAATTTGGCTGTGGAAATATGTCATAGCTGACGCTGCCGTTTTTATC 576
Db 1519330 GATTTATAGATGA 1519343

; NAME/KEY: misc feature
; LOCATION: (145171)..(145171)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (145942)..(145942)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (147197)..(147197)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (150841)..(150841)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152500)..(152500)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (152530)..(152530)

Query Match      5.0%; Score 38; DB 18; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 57;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 244 GGAATCCGAGAGAAAGGCTGACGAAAGTAGGAAAGAGGATATCAGCAAAACAGAGA 303
Db 1519210 GAATTCAAAATAATGATGGCGAGAAACCAAAAGTCAAAATCTCGAAGCGGAAGTCT 1519269

Qy 304 TTGCGAGAAAGAGCCCTTGCTGCTCATGGCGGGGATCCGAGGCGATTCATACATTTTCT 363
Db 1519270 TTACCTGATACAGTCATGCTGCTGTTGGCGGTGGCTCGAATGCAATGCTATGTTGCT 1519329

Qy 364 GATTTATAGATAA 377
Db 1519330 GATTTATAGATGA 1519343

RESULT 7
US-10-425-114-29841
; Sequence 29841, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29841
; LENGTH: 945
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMROPIC006D11_FLI
US-10-425-114-29841

Query Match      4.8%; Score 36.4; DB 16; Length 945;
Best Local Similarity 51.9%; Pred. No. 2.4;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 502 ATGATGGAAGAAGTCAGTCAGGATATGCTGAATTTGCTGTGGAATATGTCATAGCTGAG 561
Db 254 ATGCTGCAAGAGATGTACAATCAATGCGCTTCTTTAGGGTCACACTTGATTGTTGGA 313

Qy 562 CTGCCGTTTTTTATCGGAGCTCCGGATATTTAGAGGTGGAGAGACACTCTTCTGCTTAT 621
Db 314 ATGGTGTGTTCCCAAGAGAGATCCGAAATTTGCCCTCTGAATGATAGACTCTCTTTTCA 373

Qy 622 CATCGTCGTTGAAGCTGGGTGGAGAGATCAAGATCAAGTAAACA 659
Db 374 AAGGATCTGTGGCGGTTTGGGATCAATTGAGGAACA 411

RESULT 8
US-10-425-114-29880
; Sequence 29880, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
```

; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29880
; LENGTH: 3179
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMROPI021F08_FLI
US-10-425-114-29880

Query Match 4.8%; Score 36.4; DB 16; Length 3179;
Best Local Similarity 51.9%; Pred. No. 4.9;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 502 ATGATGGAAGAAGTCAGTCAGGAGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2525 ATGCTGCAAGAGATGTACAATCAATGCGCCTTCTTTAGGGTCACACTTGATTTGGTGGAA 2584
Qy 562 CTGCGGTTTTTATPCGGAGCTCCGGATATTTTAGAGGTGGAAGACACACTCTTGTCTTAT 621
Db 2585 ATGGTGTTCGCAAGAGATCCGAAATTTGCCGCTCTGAATGATAGACTCCTTGTTC 2644
Qy 622 CATCGTCCGTGGAAGCTGGGTGGAAGATCAGTAAACCA 659
Db 2645 AAGGATCTGTGGCGTTTGGGGATCAATTGAGGAACAA 2682

RESULT 9

US-10-425-114-11232
; Sequence 11232, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 11232
; LENGTH: 3285
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701043122_FLI
US-10-425-114-11232

Query Match 4.8%; Score 36.4; DB 16; Length 3285;
Best Local Similarity 51.9%; Pred. No. 5;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 502 ATGATGGAAGAAGTCAGTCAGGAGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2492 ATGCTGCAAGAGATGTACAATCAATGCGCCTTCTTTAGGGTCACACTTGATTTGGTGGAA 2551
Qy 562 CTGCGGTTTTTATPCGGAGCTCCGGATATTTTAGAGGTGGAAGACACACTCTTGTCTTAT 621
Db 2552 ATGGTGTTCGCAAGAGATCCGAAATTTGCCGCTCTGAATGATAGACTCCTTGTTC 2611
Qy 622 CATCGTCCGTGGAAGCTGGGTGGAAGATCAGTAAACCA 659
Db 2612 AAGGATCTGTGGCGTTTGGGGATCAATTGAGGAACAA 2649

RESULT 10

US-10-424-599-96841
; Sequence 96841, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 96841
; LENGTH: 3542
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58459C.1
US-10-424-599-96841

Query Match 4.8%; Score 36.4; DB 16; Length 3542;
Best Local Similarity 51.9%; Pred. No. 5.2;
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 502 ATGATGGAAGAAGTCAGTCAGGAGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAG 561
Db 2622 ATGCTGCAAGAGATGTACAATCAATGCGCCTTCTTTAGGGTCACACTTGATTTGGTGGAA 2681
Qy 562 CTGCGGTTTTTATPCGGAGCTCCGGATATTTTAGAGGTGGAAGACACACTCTTGTCTTAT 621
Db 2682 ATGGTGTTCGCAAGAGATCCGAAATTTGCCGCTCTGAATGATAGACTCCTTGTTC 2741
Qy 622 CATCGTCCGTGGAAGCTGGGTGGAAGATCAGTAAACCA 659
Db 2742 AAGGATCTGTGGCGTTTGGGGATCAATTGAGGAACAA 2779

RESULT 11

US-10-296-115-256/c
; Sequence 256, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1e1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 256
; LENGTH: 736
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-256

Query Match 4.7%; Score 35.4; DB 16; Length 736;
Best Local Similarity 50.9%; Pred. No. 4.4;
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
Qy 454 GACATGTCGTCAAGCGATAATCGGGCGTGC GGCGCGTCAGTTTGTGATGATGGAAGAA 513
Db 369 GTACAGCTGTGGAATGAGGAGCTCTTTGGGGATGGGCCCCCTTTGTAAAGTGAAGGA 310
Qy 514 GTCAGTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGCTAGCTGCCCTTTTTT 573
Db 309 GTCAGTCTGGGGACTGGAGGCCCTCTTTGGAGGATGGGGTAGCTGGGGCTTCTTTGGGG 250
Qy 574 ATCGGAGCTCCGGATATTTTAGAGGTGGAAGACACTCCTTGTCT 618
Db 249 GTTGGAAATGCTGGGGTCTTTTATAGGGGAGGGAACAGTCATAGCT 205

```
RESULT 12
US-10-424-599-6970
; Sequence 6970, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J.
; APPLICANT: Kovalic David K.
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 6970
; LENGTH: 3681
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_106301C.1
US-10-424-599-6970

Query Match      4.7%; Score 35.4; DB 16; Length 3681;
Best Local Similarity 55.2%; Pred. No. 11; Mismatches 56; Indels 0; Gaps 0;
Matches 69; Conservative 0;

QY 512 AAGTCAGTGGAGGATGCTGAATTTGGCTGTGGAATATGTCATAGCTGAGCTGCCGTTTT 571
DB 1393 AAGTGAGTTATGATGCTTTGGAGAGGATGAGCAAGTATTTCTTGACATTTGCTTGT 1452
QY 572 TTATCGGAGCTCCGGATATTTTAGAGTGGAGAGACACCTCTTGTATCATGTCCTCGT 631
DB 1453 GCTTCAAAAATATGATTTGGCAGAGTCCCAAGATATACTTCATGCTCATATGCGCAT 1512
QY 632 GGAAG 636
DB 1513 GCATG 1517

RESULT 13
US-09-814-353-21372
; Sequence 21372, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21372
; LENGTH: 4061
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-21372
```

```
Query Match      4.7%; Score 35.4; DB 10; Length 4061;
Best Local Similarity 50.9%; Pred. No. 12; Mismatches 81; Indels 0; Gaps 0;
Matches 84; Conservative 0;

QY 454 GACATGCTCGTGAAGGATAATCGGGCGTGGCGGGCGCTCAGTTTGATGATGGAAGAA 513
DB 3718 GTCACAGCTGGTGGAAATGAGGAGCTCTTTGGGGGATGGGGCCCTTTTGAAGTGAAGGA 3777
QY 514 GTCAGTGAAGATATGCTGAATTTGGCTGTGCAATATGTCATAGCTGAGCTGCCGTTTTTT 573
DB 3778 GTCACCTGCTGGGGACTTGGAGGCTCTTTGAGAGATGGGTAGCTGGGGCTTCTTTGGGG 3837
QY 574 ATCGGAGCTCCGGATATTTTAGAGTGGGAAGAGACACTCCTTGCT 618
DB 3838 GTTGAATTGCTGGGGTCTTTTATAGGGAGGGAACAGTCATAGCT 3882

RESULT 14
US-09-764-891-7281/c
; Sequence 7281, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7281
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7281

Query Match      4.7%; Score 35.4; DB 10; Length 13555;
Best Local Similarity 50.9%; Pred. No. 23; Mismatches 81; Indels 0; Gaps 0;
Matches 84; Conservative 0;

QY 454 GACATGCTCGTGAAGGATAATCGGGCGTGGCGGGCGCTCAGTTTGATGATGGAAGAA 513
DB 9130 GTCACAGCTGGTGGAAATGAGGAGCTCTTTGGGGGATGGGGCCCTTTTGAAGTGAAGGA 9071
QY 514 GTCAGTGAAGATATGCTGAATTTGGCTGTGCAATATGTCATAGCTGAGCTGCCGTTTTTT 573
DB 9070 GTCACCTGCTGGGGACTTGGAGGCTCTTTGAGAGATGGGTAGCTGGGGCTTCTTTGGGG 9011
QY 574 ATCGGAGCTCCGGATATTTTAGAGTGGGAAGAGACACTCCTTGCT 618
DB 9010 GTTGAATTGCTGGGGTCTTTTATAGGGAGGGAACAGTCATAGCT 8966

RESULT 15
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1163020
```



```
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

Query Match          4.78; Score 35.4; DB 16; Length 1163020;
Best Local Similarity 46.28; Pred. No. 2.9e+02;
Matches 117; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 195 TTCAAGTTTACTTGCAGGCGCATGAGCGGCTAATCTTCTAGAAAGCGCTTGGAACTCCGAG 254
Db 892538 TTTGATGTTGCAAGAAGGGAAGTTCTTCAACGCGTATCGTATCGTATGCGCATTCGTC 892597

Qy 255 AGGAAAGGCTGAACGAAAGTAAGGAAAGAGGTATACGAAACAGGAGATTTCGAGAAAG 314
Db 892598 GGCAAGTACGCAAGAACAGAAAGCGTTACAACTACCAAGAAACAGCAATGTAATGAAAT 892657

Qy 315 AGCCCTTGTGGCTCATGCGGGGATCCGAAGCGGATTCATACATTTTCTGATTTATAGA 374
Db 892658 TGAACGGATTTCGTTACGCGGACCGGTGTAACCAATTTTATGAAGTTGCCGCCATTCAGA 892717

Qy 375 TAACAAAGCCTACCAGCTGTTTGAGACAAAGAGTTGAACATGTCATTTTTCGAGCAGCCTCA 434
Db 892718 AAAATTGCTTCACGTGTTACAAAGGAGACATTTATGATTCCTTTATAAAGCGATTGA 892777

Qy 435 TTTTCGACATGCT 447
Db 892778 ATTAAACTTGGT 892790
```

Search completed: December 10, 2004, 21:11:32
Job time : 469.285 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:46:32 ; Search time 159.553 Seconds
(without alignments)
7233.283 Million cell.updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 gtaagatgagatgccggaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1633475 seqs, 766331085 residues

Total number of hits satisfying chosen parameters: 3266950

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA New.*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US11_NEW_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	346.2	46.0	4222334	1	PCT-US04-37204-1
2	345.6	45.9	750	1	PCT-US04-37204-3406
3	43.2	5.7	1101	1	PCT-US03-31318-9046
4	43.2	5.7	1457619	1	PCT-US03-31318-8739
5	38	5.0	1194	1	PCT-US02-09107B-22329
6	38	5.0	1830121	6	US-10-981-687-1
7	36.2	4.8	141753	6	US-10-746-294A-28
8	35.4	4.7	577	1	PCT-US04-35274-53829
9	35.4	4.7	600	1	PCT-US04-35274-53828
10	35.4	4.7	51879	6	US-10-990-328-94645
11	35	4.6	32943	6	US-10-990-328-94732
12	34.8	4.6	408	6	US-10-275-323A-13
13	34.4	4.6	429	6	US-10-948-737-5526
14	34.2	4.5	1446	1	PCT-US02-09107B-18304
15	34.2	4.5	499008	6	US-10-990-328-96831
16	34	4.5	198617	6	US-10-990-328-97086
17	33.8	4.5	963	1	PCT-US02-09107B-18554
18	33.8	4.5	1660	6	US-10-801-292-5
19	33.8	4.5	181323	6	US-10-990-328-93896
20	33.6	4.5	44322	6	US-10-990-328-93477
21	33.4	4.4	501	1	PCT-US02-09107B-31088
22	33.2	4.4	373600	6	US-10-990-328-93686
23	33	4.4	200	1	PCT-US03-31318-2361
24	33	4.4	28892	6	US-10-990-328-93860
25	32.8	4.4	44186	6	US-10-990-328-97821

ALIGNMENTS

RESULT 1

PCT-US04-37204-1
; Sequence 1, Application PC/TUS0437204
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Bacillus Licheniformis Chromosome
; FILE REFERENCE: 10588.204-WO
; CURRENT APPLICATION NUMBER: PCT/US04/37204
; CURRENT FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: 60/535,988
; PRIOR FILING DATE: 2004-01-09
; PRIOR APPLICATION NUMBER: 60/561,059
; PRIOR FILING DATE: 2004-04-08
; PRIOR APPLICATION NUMBER: 60/572,403
; PRIOR FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 8395
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4222334
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
PCT-US04-37204-1

Query Match	46.0%	Score 346.2;	DB 1;	Length 4222334;
Best Local Similarity	66.3%;	Pred. No. 1.4e-103;		
Matches 498;	Conservative 0;	Mismatches 253;	Indels 0;	Gaps 0;
Qy	3	GAATGATGATACCGGAAATG	TATCGGAAGAGGTCTGTG	CGATTTTATTCTCTAGGCAATT 62
Db	3411675	GAAAATCAATGCAGAGCGTT	ATAATGGAGAGCAACACCA	CGCTATTCAAAACCGGAAACTCT 3411734
Qy	63	AACAGAAAACCTGCAGAGAA	ATATTTGAAACGGCGCAGGC	ATGTTTTGTGGGGATCAGCCC 122
Db	3411735	TACCCAAAACCTGCAATGAA	ATATTTAAACGCGACAGCGC	ATGTTCTCGTGGCATCAGCCC 3411794
Qy	123	ATTTTAACAGCAGCGTTTTTC	AGAGGATTTATTTACAGATT	AAATTTGGATGGCGGAAAGCTCA 182
Db	3411795	GTTTAAACAGCCGATTTTTC	GAGATTAATTCATCGGCTT	ATCGCTGGGCGCTCGGTGA 3411854
Qy	183	ATTTTAAAGCGTTTCAGTTT	TACTTTCAGGGCATAGCGG	CGCTTAATCTTCTAGAAAGCGCT 242
Db	3411855	GTTTTCAGATGTATCCGTCT	TTTGGCGGAAAGGAGCTGC	CCAACTTCTCGAAGCGCT 3411914
Qy	243	TGGAATCCGAGAGGAAAGCT	GAAACGAAAGTAAAGGAA	AGAGTATCAAGAAACAGGAG 302
Db	3411915	CGGACCCCATCATGCGAAG	CGCGAAACGAAAGTCAG	AAAGAGATCTCGCGGAACCGGAG 3411974

QY 303 ATTTCAGAAAGACCCCTTGTGGCTCATGTCGGGGATCCGAAGCGATTTCATACATTTTC 362
DB 3411975 ATTTCGTAAAGCGCTTGAAGCGCATGCGGAAATCCCGAGACATCCATACATTTTC 3412034
QY 363 TGATTTTATAGATAACAAAGCCCTACCAAGCTGTGTTGAGACAAAGAGTTGAAACATGCATTTT 422
DB 3412035 CGATTTTCGGAACACAGACCGCATACCGGAATTTTGGGATGGAAGTCGAAGCTGCCCTTTT 3412094
QY 423 TGAGCAGCTCATTTTCGACATGCTTGTGTCATGTCCTGTCGATGTCCTGTCGAGCGATTAATCGGGC 482
DB 3412095 CGACCAAGCAGCTTTTTCGCAATGCCCTGTCGAGATGTCGATGCCGCTATCTCCGAGC 3412154
QY 483 TCGCGGGGGCTCAGTTTGTATGATGGAAGAGTCAGTGAGATATGCTGAAATTTGGCTGT 542
DB 3412155 GGCGGGGGCCTCGGATGATGTCGTGGAAGTCAGCGCAGACATGCTGAGCTGCTGT 3412214
QY 543 GGAATATGTCATAGCTGAGCTGCCGTTTTTTTATCGGAGCTCCCGATATTTTAGAGTGA 602
DB 3412215 TGAATAGCTCATGCTGAACTTCCGTTTTTTCATCGCGCCCTGATATTTTAGCGCTGA 3412274
QY 603 AGAGACACTCTGCTTATCATGTCGTCGTCGGAAGCTGGGTGAGAAGATCAGTAACCATGA 662
DB 3412275 AGAGACGCTTCTGCTTATCAGCGCCCATGGAAGCTCGCGAACAGATCTCCCGTAATGA 3412334
QY 663 ATTTTCTATTGATGCGCGCAATCAAGGTATCTCATTTGACAGGAATGCGCAGAT 722
DB 3412335 ATTGCGCTCAAAATGGCGCGCAATCAAGATATCTCATGTTCCGACCGGACGAAAG 3412394
QY 723 GCTTCTGAGAAACGGATCATCTCTGAAGGA 753
DB 3412395 GGTGGAATCTAAGCATGACGAGGAACGA 3412425

RESULT 2

PCT-US04-37204-3406
; Sequence 3406, Application PC/TUS0437204
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; APPLICANT: NOVOZYMES A/S
; TITLE OF INVENTION: Bacillus Licheniformis Chromosome
; FILE REFERENCE: 10588.204-WO
; CURRENT APPLICATION NUMBER: PCT/US04/37204
; CURRENT FILING DATE: 2004-11-05
; PRIOR FILING DATE: 2004-01-09
; PRIOR FILING DATE: 2004-01-09
; PRIOR FILING DATE: 2004-04-08
; PRIOR FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 8395
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3406
; LENGTH: 750
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
PCT-US04-37204-3406

Query Match 45.9%; Score 345.6; DB 1; Length 750;
Best Local Similarity 66.5%; Pred. No. 1.7e-105;
Matches 495; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 10 ATGACCGAATGGTAACGGAAGAGTCTGTGATTTTATTGCTGAGGCATTAACAGAA 69
DB 1 ATGACAGAGCTTAATATGGAGAGCAACACAGCATATTCAAAACCGAAATCTTTACCCAA 60
QY 70 AACTGCAGAGAAATATTGTAACGCGCAGGAGCATGTTTGTGGGATCAGCCCATTTAAC 129
DB 61 AACTGCAATGAAATATTAAACGAGAGCGCAUGTTCTGTCGCAATCAGCCGTTTAAAC 120
QY 130 AGCAGGTTTTTCAGAGGATTAATTATACAGATTAATGGAATGGGCAAGCTCAATTTAA 189
DB 121 AGCGAATTTTCGGAAGATTAATATCATCGGCTTATCGCTGGGCGTCCGCTGAGTTTCAG 180

QY 190 AGCGTTTCAGTTTACTTTCAGAGGCATGAGCGGCTAAATCTTCTAGAAGCGCTTGGAACT 249
DB 181 AGTGATCCGTGCTTTTTCGCGGGAAGAGCTGCCAACCCTTCTCGAAGCGCTCGCAC 240
QY 250 CCGAGAGAAAGGCTGAAACGAAAGTAAAGAAAGAGGTATCAGAAACAGGAGATTGCA 309
DB 241 CCACATGGGAAGCGCCGAAACGAAAGTCAAGAAAGTCTCGCGAAACCGGAGATTGCT 300
QY 310 GAAAGAGCCCTTGTGCTCATGCGGGGATCCGAAGCGATTTCATACATTTTCTGATTTT 369
DB 301 GAAAGAGCCCTTGAAGCGCATGCGGAAATCCCGAGACATCCATACATTTTCCGATTT 360
QY 370 ATAGATAACAAAGCCCTACCAAGCTGTTGAGACAAAGAGTTGAAACATGCATTTTGAAGCAG 429
DB 361 GCGAACCAAGCCGATACCGGAATTTTTCGGATGGAAGTCGAAGCTGCCCTTTTCGACCA 420
QY 430 CTTATTTTCGACATGCTTGTGTCATGTCCTGTAAGCGGATATTCGGGCGTTCGCGCG 489
DB 421 AGCAATTTTCGCAATGCTTCCCTGAGATGTCGATCGCGCTATCTCCGACGCGGCGCG 480
QY 490 GCGCTCAGTTTGTATGATGGAAGAGTCAGTGAGGATATGCTGAATTTGGCTGTGGAATAT 549
DB 481 GGCATCTCGATGATGTCGTGGAAGTCAGCGCAGACATGCTGGAGCTGGCTGTTGAATAC 540
QY 550 GTCATAGCTGAGCTGCGGTTTTTTATCGGAGCTCCGATATTTTAGAGGTGGAAGACGA 609
DB 541 GTCATGCTGAACTTCCGTTTTTTCATCGCGCCCTGATATTTTAGGCGTTCGAAGACG 600
QY 610 CTCCTTGTTCATCCTCGTGGAAAGCTGGGTGGAAGATCAGTAAACCATGAATTTTCT 669
DB 601 CTTCTTGTTCATCAGCGGCTCGGAAGCTCGCGAAGACAGATCTCCGTAATGAATTTGCC 660
QY 670 ATTTGTATGCGCGCAATCAAGGTATCTCATTTGACAGAAATGCGCAGATGCTTCT 729
DB 661 GTCAAAATCGCGCAATCAAGGATATCTCATGTTTCCGAAGCGGACGAAGGTTGAA 720
QY 730 GAGAAACGGATCACATCTGAAGGA 753
DB 721 TCTAAAAGCATGCGAGGGAACGA 744

RESULT 3

PCT-US03-31318-9046/c
; Sequence 9046, Application PC/TUS0331318
; GENERAL INFORMATION:
; APPLICANT: Regence of the University of Minnesota
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128WO1
; CURRENT APPLICATION NUMBER: PCT/US03/31318
; CURRENT FILING DATE: 2003-10-01
; PRIOR FILING DATE: 2003-10-01
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9046
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
PCT-US03-31318-9046

Query Match 5.7%; Score 43.2; DB 1; Length 1101;
Best Local Similarity 60.0%; Pred. No. 0.00088;
Matches 72; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 350 TTCATACATTTCTGATTTTATAGATAACAAAGCCTACCAAGCTGTTGAGACAAGAGTTG 409
DB 561 TTCATTTTCTCATTAATGATGAGGTTTCCACTATTTAAGTGAAGAAAG 502
QY 410 AACATGATTTTGTAGAGCCCTCATTTTCGACATGCTTGTGGAATGCTCTCTGAAG 469
DB 501 AAGATTCAATTTCTCAACGCCCATTTAGCGCTTTTATTTATTTTCTGCTATTTTGAAG 442

```
RESULT 4
PCT-US03-31318-8739
; Sequence 8739, Application PC/TUS0331318
; GENERAL INFORMATION:
; APPLICANT: Regents of the University of Minnesota
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128W01
; CURRENT APPLICATION NUMBER: PCT/US03/31318
; CURRENT FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 8739
; LENGTH: 1457619
; TYPE: DNA
; ORGANISM: Lawsonia intracellularis
PCT-US03-31318-8739

Query Match      5.7%; Score 43.2; DB 1; Length 1457619;
Best Local Similarity 60.0%; Pred. No. 0.051;
Matches 72; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 350 TTCATACATTTCTCGATTATTTATAGATAACAAAGCCTACCGCTGTTGAGACAAAGATTG 409
Db 184520 TTCATTTTCTCTCATTAATAATAGATGCGAGGTTTCCACTATTAAAGTGAAGAAG 184579

Qy 410 AACATGCATTTTGTGACGCCCTCATTTTCGACATGCTTCTTGGACATCTCTCGTGAAG 469
Db 184580 AAGATTCATTCTTCAACAGCCACTTTAGCGTCTTTATTATTATTTTTCGTATTTTGAAG 184639

RESULT 5
PCT-US02-09107B-22329
; Sequence 22329, Application PC/TUS0209107B
; GENERAL INFORMATION:
; APPLICANT: Elittra Pharmaceuticals Inc.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034VPC
; CURRENT APPLICATION NUMBER: PCT/US02/09107B
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/342,923
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 10/072,851
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/362,699
; PRIOR FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22329
; LENGTH: 11194
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
PCT-US02-09107B-22329

Query Match      5.0%; Score 38; DB 1; Length 11194;
Best Local Similarity 55.2%; Pred. No. 0.052;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 244 GGAACCTCGAGAGAAAGCGTGAACGAAAGTAAGGAAAGGATATCCGAAACAGAGA 303
Db 610 GAATTTCAAAAATGATTGGCGAAGAAACCAAAACGTCAAAATTCGAAACGGAAGTCTG 669

Qy 304 TTTCCAGAAAGCCCTTGTGGCTCATGGCGGATCCGAAGCGATTACATATTTCT 363
Db 670 TTACCTGATACAGTCATTGCTGCTGTTGGCGGTGCGTCCGAATGCAATGGTATGTTGCT 729
```

```
Qy 364 GATTTATAGATAA 377
Db 730 GATTTATGATGA 743

RESULT 6
US-10-981-687-1
; Sequence 1, Application US/10981687
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
; FILE REFERENCE: PB186P2C1D12
; CURRENT APPLICATION NUMBER: US 10/981,687
; CURRENT FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: US 10/158,865
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a,t,c, or g
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
```

```
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc feature
```

Query Match 5.0%; Score 38; DB 6; Length 1830121;
Best Local Similarity 55.2%; Pred. No. 3.2;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy 244 GGAACCTCCGAGAGGCTGAACGAAAGTAAGGAAAGAGGTATCCGAAACAGGAGA 303
Db 1519210 GAATTTCAAAATGATTGCGAAGAACCAACAGTCAAATTCGAAACGGAAGTCT 1519269

Qy 304 TTTCGAGAAAGAGCCCTTGTGGCTCATGTGGCGGGATCCGAAAGCGATTACATTTTCT 363
Db 1519270 TTACCTGATACAGTCATTGCTGTGGCGGTGGCTCGAATGCAATTGGTATGTTTGTCT 1519329
```

```
Qy 364 GATTTTATAGATAA 377
|||||
Db 1519330 GATTTTATGATGA 1519343

RESULT 7
US-10-746-294A-28/c
; Sequence 28, Application US/10746294A
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)E
; CURRENT FILING DATE: 2003-12-29
; PRIOR APPLICATION NUMBER: US/10746,294A
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 09/692,412
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 09/803,736
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 124
; SEQ ID NO 28
; LENGTH: 141753
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(141753)
; OTHER INFORMATION: unsure at all n locations
US-10-746-294A-28

Query Match 4.8%; Score 36.2; DB 6; Length 141753;
Best Local Similarity 46.9%; Pred. No. 3.1; Indels 0; Gaps 0;
Matches 113; Conservative 0; Mismatches 128;

Qy 139 TCAGAGGATTATATTTACAGATTAAATGGATGGCGGAAAGCTCAATTTAAAGCGTTTCA 198
Db 124511 TCGTTGAACTAATCTCAGTGACAGAGTTTCAGAGAGATTCATGTTTCAAGATGATGCA 124452

Qy 199 GTTTTACTTCAGGCGATAGGCGGCTAATCTTCTAGAACGGCTTGGAACTCCGAGAGGA 258
Db 124451 ATTTTAAAGGATACACAAAAACGAAGAAGATAGAAACAGTTTCTATTGGTAGTGAG 124392

Qy 259 AAGCTGAAGCAAGTAAGGAAGAGGTATCAGCAACAGGAGATTTGCAGAAAGAGCC 318
Db 124391 AAAGTTGTGAGAGAGAGGTGAAGCTGAGAAAGACGGTGGAGGATTTTGGAAAAACGCT 124332

Qy 319 CTGTGGCTCATGGCGGGGATCCGAAGCGATTTCATACATTTTCTGATTTTATAGATAAC 378
Db 124331 GTTGTAGGTATTGTGTTCTCCGTTGGACAGTGCGCGGATGATCTCGGAGTAAG 124272

Qy 379 A 379
Db 124271 A 124271

RESULT 8
PCT-US04-35274-53829/c
; Sequence 53829, Application PC/TUS0435274
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER
; FILE REFERENCE: MM1110-1W0
; CURRENT APPLICATION NUMBER: PCT/US04/35274
```

```
; CURRENT FILING DATE: 2004-11-10
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 53829
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Chicken 19866894305695_3
PCT-US04-35274-53829

Query Match 4.7%; Score 35.4; DB 1; Length 577;
Best Local Similarity 52.3%; Pred. No. 0.26; Indels 0; Gaps 0;
Matches 78; Conservative 0; Mismatches 71;

Qy 4 AATGAGATGACCGGAATGGTAACGGAAGAAGGTCTGTGCATTTTATTTGCTGAGGCATTA 63
Db 449 AATGCCAGCACTTCAACAAAACCCAGCAGGATGCATTAGTACCACATCAGCCAGAACTGA 390

Qy 64 ACAGAAAACTGCAGAGAAATATTTGAACGGCGGAGGATGTTTGGTGGGATCAGCCCA 123
Db 389 ACCGTATCTGCCAAGCAAACTTTGGAAGGAGTCAGTCAAGTTTGTGTAAGAGCCTTCAA 330

Qy 124 TTTAACAGCAGGTTTTCAGAGGATTATAT 152
Db 329 GTCAACGCGACCATTAACAATGATGACAT 301

RESULT 9
PCT-US04-35274-53828/c
; Sequence 53828, Application PC/TUS0435274
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER
; FILE REFERENCE: MM1110-1W0
; CURRENT APPLICATION NUMBER: PCT/US04/35274
; CURRENT FILING DATE: 2004-11-10
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 53828
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894305695_2
PCT-US04-35274-53828

Query Match 4.7%; Score 35.4; DB 1; Length 600;
Best Local Similarity 52.3%; Pred. No. 0.26; Indels 0; Gaps 0;
Matches 78; Conservative 0; Mismatches 71;

Qy 4 AATGAGATGACCGGAATGGTAACGGAAGAAGGTCTGTGCATTTTATTTGCTGAGGCATTA 63
Db 558 AATGCCAGCACTTCAACAAAACCCAGCAGGATGCATTAGTACCACATCAGCCAGAACTGA 499

Qy 64 ACAGAAAACTGCAGAGAAATATTTGAACGGCGGAGGATGTTTGGTGGGATCAGCCCA 123
Db 498 ACCGTATCTGCCAAGCAAACTTTGGAAGGAGTCAGTCAAGTTTGTGTAAGAGCCTTCAA 439

Qy 124 TTTAACAGCAGGTTTTCAGAGGATTATAT 152
Db 438 GTCAACGCGACCATTAACAATGATGACAT 410

RESULT 10
US-10-990-328-94645/c
; Sequence 94645, Application US/10990328
```

```
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 94645
; LENGTH: 51879
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-990-328-94645

Query Match          4.7%; Score 35.4; DB 6; Length 51879;
Best Local Similarity 50.9%; Pred. No. 3.2;
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 454 GACATGCTCTGTAAGCGAATAATCGGCGGTGCGGGGGCTCAGTTTGATGATGGAAGAA 513
Db 41590 GTCACAGCTGTGGAATGAGGAGCTCTTTGGGGGATGGGGCCCTTTGTAAAGTGAAGGA 41531

QY 514 GTCAGTGAAGATATGCTGAATTTGGCTGTGGAATATGCTAGCTAGAGTGCCTGTTT 573
Db 41530 GTCACGCTGGGAGCTGGAGGCTCTTTGGAGGATGGGTAGCTGGGGCTTCTTTGGGG 41471

QY 574 ATCGGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCTCTTGT 618
Db 41470 GTTGGAAATGCTGGGTCTTTTATGGGAGGGAACAGTCATAGCT 41426

RESULT 11
US-10-990-328-94732
; Sequence 94732, Application US/10990328
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 94732
; LENGTH: 32943
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(32943)
; OTHER INFORMATION: n = A,T,C or G
US-10-990-328-94732

Query Match          4.6%; Score 35; DB 6; Length 32943;
Best Local Similarity 49.7%; Pred. No. 3.4;
Matches 89; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 30 AAGAAGTCTGTGCATTTTATTGCTGAGGCATTAAACAGAAACATGCAGAGAAATATTGA 89
Db 21106 AGGAAGTACTTACATTATCTCTATTAAACAGAGAAATATGTGAAGATTTTGT 21165

QY 90 ACGGCGCAGGCATGTTTTGGTGGGATCAGCCCATTTTAAACAGCAGGTTTTTTCAGAGGATTA 149
Db 21166 CCATGTCACACAGTTAGTGGTGGCAGAACTGCAGATTCAAAACCCAGTGTTCAGAGACTGC 21225

QY 150 TATTACAGATTAAATGATGGGGAAGCTCAATTTAAAGCGTTTCAGTTTACTTT 208
Db 21226 TTTTAAACCACTACATTCTTCTGTCCCTCTAATCATGAATTAATGCAACTTTTACTTTG 21284
```

```
RESULT 12
US-10-275-323A-13/c
; Sequence 13, Application US/10275323A
; GENERAL INFORMATION:
; APPLICANT: DONNE-GOUSSE, CAROLE
; APPLICANT: LAUDET, VINCENT
; APPLICANT: HANNI, CATHERINE
; TITLE OF INVENTION: METHOD OF DETECTING AND IDENTIFYING THE PRESENCE OF BIOLOGICAL
; TITLE OF INVENTION: MATERIALS DERIVED FROM BIRDS, AND OLIGONUCLEOTIDES FOR CARRYING
; TITLE OF INVENTION: IT OUT
; FILE REFERENCE: 0508-1052
; CURRENT APPLICATION NUMBER: US/10/275,323A
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01279
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: FR 00/05850
; PRIOR FILING DATE: 2000-05-09
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA fragment amplified
; OTHER INFORMATION: from the genome of birds
; NAME/KEY: modified_base
; LOCATION: (43)..(43)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (57)..(57)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (71)..(71)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (134)..(134)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (212)..(212)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (214)..(214)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (217)..(217)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (226)..(227)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (236)..(236)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (245)..(245)
; OTHER INFORMATION: a, c, g, or t
; NAME/KEY: modified_base
; LOCATION: (253)..(253)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (269)..(269)
```


Job time : 182.553 secs

```
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/342,923
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 10/072,851
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: 60/362,699
; PRIOR FILING DATE: 2002-03-06
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18304
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
PCT-US02-09107B-18304

Query Match      4.5%; Score 34.2; DB 1; Length 1446;
Best Local Similarity 49.7%; Pred.No.1.1;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 206 TTCGAGGGCATGAGGGCGCTAATCTTCTAGAGCGCTTGGAACCTCCGAGAGGAAAGGCTG 265
Db 92 TTGCAGTTTTACAGAGCCCTGATCGTGATGTTTGCATGTAAAAATCGCAGATGAGCTT 151

Qy 266 AACGAAAAGTAAGAAAGAGGATATCACGAAACAGGAGATTTCGAGAAAGAGCCCTTGTGG 325
Db 152 ATCGCATAGGACGGATGCTATAAGAGGATATTTAGATGTTGCTCGTATTGTCGAGATTG 211

Qy 326 CTGATGGCGGGATCCGAAGCGATCATACATTTTCTGATTTTATAGATAACAA 380
Db 212 CTAAGGCTTGTGTGCTGATGCGATTCATCCTGTTATGGATTTTAAAGTGAGAA 266

RESULT 15
US-10-990-328-96831/c
; Sequence 96831, Application US/10990328
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; FILE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001495
; CURRENT APPLICATION NUMBER: US/10/990,328
; CURRENT FILING DATE: 2004-11-17
; NUMBER OF SEQ ID NOS: 558824
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 96831
; LENGTH: 499008
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(499008)
; OTHER INFORMATION: n = A,T,C or G
US-10-990-328-96831

Query Match      4.5%; Score 34.2; DB 6; Length 499008;
Best Local Similarity 52.4%; Pred.No.29;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 158 GATTAATTGGATGGCGGAAGCTCAATTTAAAGCGTTTCAGTTTTCCTTGCAGGGCATG 217
Db 129915 GCTTGGGTGGAGTGGGTGAGTACATAGAAAGGGGCGCAAGAAACAAAGAGAGGTAGTG 129856

Qy 218 AGCGGGCTAATCTTCTAGAGCGCTTGGAACCTCCGAGAGGAAAGGCTGAAACGAAAGTAA 277
Db 129855 TGGAAAGATAAGAGGTTAGACAAAGCAGAGAACCCCTTAGCGAGAGGGATCAGCATATAGAA 129796

Qy 278 GGAAGAGGTTATCAGAAACAGG 300
Db 129795 GGAATGAAGTATGAGGAGCATG 129773

Search completed: December 10, 2004, 20:51:35
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 13:18:28 ; Search time 2775.16 Seconds
(without alignments)
9887.411 Million cell updates/sec

Title: US-10-627-124-7
Perfect score: 753
Sequence: 1 ggaatgagatgaccggaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	43.2	5.7	585	8	BZ484107
C 2	43.2	5.7	818	8	BH431961
C 3	43.2	5.7	825	8	BH555688
C 4	43.2	5.7	844	8	BZ429521
C 5	43.2	5.7	916	9	CNS005MM
C 6	40	5.3	1013	9	CL053107
C 7	40	5.3	1101	9	CNS0039G
C 8	39.6	5.3	414	8	AZ001661
C 9	39.6	5.3	498	8	AZ289765
C 10	39.6	5.3	552	8	AZ511287
C 11	39.6	5.3	717	9	AG517041
C 12	39.4	5.2	309	7	CN215521
C 13	39	5.2	926	7	CR442585
C 14	39	5.2	1101	9	CNS017KX
C 15	38.4	5.1	1021	9	CNS00DLS
C 16	38.4	5.1	1103	9	CL082028
C 17	38.2	5.1	1098	8	CC187205
C 18	38	5.0	555	2	AW705941
C 19	37.8	5.0	802	5	BX723518
C 20	37.6	5.0	800	8	BZ470290
C 21	37.6	5.0	1101	9	CNS0100X
C 22	37.4	5.0	491	1	AA516933
C 23	37.4	5.0	505	8	AQ721360
C 24	37.4	5.0	543	6	CB929916

25	37.4	5.0	589	2	AW032477	AW032477 EST276036
C 26	37.4	5.0	703	8	BH936464	BH936464 OG960803
C 27	37.2	4.9	620	5	BQ527323	BQ527323 NISC_no21
C 28	37.2	4.9	1101	9	CNS00LOO	AL068607 Drosophila
C 29	37	4.9	466	9	CG470635	CG470635 ZMMBB5027
C 30	37	4.9	766	8	AZ600588	AZ600588 LM0418B05
C 31	36.8	4.9	307	9	CL571383	CL571383 OB_BA001
C 32	36.8	4.9	414	8	AQ084406	AQ084406 HS_2198B
C 33	36.8	4.9	480	9	CL602544	CL602544 OB_BA007
C 34	36.8	4.9	531	9	CL582438	CL582438 OB_BA004
C 35	36.8	4.9	570	4	B1785498	B1785498 sal1f11
C 36	36.8	4.9	733	8	BZ028160	BZ028160 oes78808
C 37	36.8	4.9	950	5	EX442989	EX442989 BX442989
C 38	36.8	4.9	1070	8	BH655582	BH655582 BOMNC81TR
C 39	36.6	4.9	676	7	CO090490	CO090490 GR_Ea101
C 40	36.4	4.8	387	5	BU050997	BU050997 1111037B0
C 41	36.4	4.8	494	2	BE030001	BE030001 sr47f05.Y
C 42	36.4	4.8	553	4	B1973129	B1973129 saf40c05
C 43	36.4	4.8	559	1	AV826780	AV826780 AV826780
C 44	36.4	4.8	587	9	CC962904	CC962904 BOB393TF
C 45	36.4	4.8	592	4	B1974352	B1974352 sai99e11

ALIGNMENTS

RESULT 1
LOCUS BZ484107/c 585 bp DNA linear GSS 13-DEC-2002
DEFINITION BZ484107 BO_1.6_2_KB tot Brassica oleracea genomic clone BOMKL22,
genomic survey sequence.
ACCESSION BZ484107
VERSION BZ484107.1 GI:26786505
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
REFERENCE 1 (bases 1 to 585)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOMKL22TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
source Location/Qualifiers
1..585
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOMKL22"
/clone_lib="BO_1.6_2_KB_tot"
/notes="Vector: pHO51; Site 1: BstXI; 1.6-2 kb sheared total DNA inserted into pHO51 using BstXI linkers"

ORIGIN
Query Match 5.7%; Score 43.2; DB 8; Length 585;
Best Local Similarity 50.0%; Pred. No. 0.23;
Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Qy 103 GTTTTGGTGGGATCCGCCATTTACAGCAGGTTTTCAGAGGATTATTATTCAGATTATTA 162
Db 504 GTTGTGTTTCAGGAACATGCTTCACCCAGAAATCATCCAAAGGAGAAAGCAGAGATGC 445

```

QY 163 ATTGGATGGCGAAGCTCAATTTAAAGCGTTTCAGTTTACTTGCAGGGCATGAGCG 222
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 AGTGAAGAGCTTCAGCTGACTCTCAAGCGTTGAAGATGGAGATGAAGAG 385

QY 223 GCTAATCTTCTAGAAGCGCTTGGAACTCCAGAGAAAGCTCAACGAAAGTAAGGAAA 282
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 GGAACATGTGCTGCAGATGAAGAAGCAAAAGGAAGAGCTGATGAAGAAGCTGTGTAGTAA 325

QY 283 GAGGTATCAGAACAGGAGATTTCAGAAAGAGCC 318
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 ANAGAGCGAGAAAACCTCCAGAGAGAGGCAAAAGAGAC 289

RESULT 2
BH431961
LOCUS BOH4111TF BOHA Brassica oleracea genomic clone BOH4111, genomic
DEFINITION survey sequence.
ACCESSION BH431961
VERSION BH431961.1 GI:17617682
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 818)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOH4111TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
  1..818
  /organism="Brassica oleracea"
  /mol_type="genomic DNA"
  /strain="TO1000DH3"
  /db_xref="taxon:3712"
  /clone="BOH4111"
  /clone_lib="BOHA"
  /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
  genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
Query Match 5.7%; Score 43.2; DB 8; Length 818;
Best Local Similarity 50.0%; Pred. No. 0.25;
Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 103 GTTTTGGTGGGATCAGCCCATTTTAAACAGCAGGTTTTCAGAGGATTTATTTACAGATTA 162
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 GTTGTGTCAGCAACATGCTTCACACCAGATCATCCAAAGGAGAGAAAGCAGAAAGATGC 135

QY 163 ATTGGATGGCGAAGCTCAATTTAAAGCGTTTCAGTTTACTTGCAGGGCATGAGCG 222
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 AGTGAAGAGCTTCAGCTGACTCTCAAGCGCTTGAAGATGGAGATGAAGAG 195

QY 223 GCTAATCTTCTAGAAGCGCTTGGAACTCCAGAGAAAGCTCAACGAAAGTAAGGAAA 282
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 GGAACATGTGCTGCAGATGAAGAAGCAAAAGGAGAGCTGATGAAGAAGCTGTGTAGTAA 255

QY 283 GAGGTATCAGAACAGGAGATTTCAGAAAGAGCC 318
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 ANAGAGCGAGAAAACCTCCAGAGAGAGGCAAAAGAGAC 291

```

```

RESULT 3
BH555688
LOCUS BOHTP71TF BOHT Brassica oleracea genomic clone BOHTP71, genomic
DEFINITION survey sequence.
ACCESSION BH555688
VERSION BH555688.1 GI:17807468
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 825)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
  1..825
  /organism="Brassica oleracea"
  /mol_type="genomic DNA"
  /strain="TO1000DH3"
  /db_xref="taxon:3712"
  /clone="BOHTP71"
  /clone_lib="BOHT"
  /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
  genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN
Query Match 5.7%; Score 43.2; DB 8; Length 825;
Best Local Similarity 50.0%; Pred. No. 0.25;
Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 103 GTTTTGGTGGGATCAGCCCATTTTAAACAGCAGGTTTTCAGAGGATTTATTTACAGATTA 162
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 GTTGTGTCAGCAACATGCTTCACACCAGATCATCCAAAGGAGAGAAAGCAGAAAGATGC 310

QY 163 ATTGGATGGCGAAGCTCAATTTAAAGCGTTTTCAGTTTACTTGCAGGGCATGAGCG 222
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 AGTGAAGAGCTTCAGCTGACTCTCAAGCGCTTGAAGATGGAGATGAAGAG 370

QY 223 GCTAATCTTCTAGAAGCGCTTGGAACTCCAGAGAGAAAGCTCAACGAAAGTAAGGAAA 282
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 GGAACATGTGCTGCAGATGAAGAAGCAAAAGGAGAGCTGATGAAGAAGCTGTGTAGTAA 430

QY 283 GAGGTATCAGAAACAGGAGATTTCAGAAAGAGCC 318
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 AAAGCGCAGAAAACCTCCAGAGAGAGGCAAAAGAGAC 466

RESULT 4
BZ429521
LOCUS BOOY611TR BO 1.6_2_KB tot Brassica oleracea genomic clone BOOY61,
DEFINITION genomic survey sequence.
ACCESSION BZ429521
VERSION BZ429521.1 GI:26673645
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 844)

```

```

high quality sequence stop: 320:
Location/Qualifiers
1. .1013
/organism="Xenopus tropicalis"

```


QY 373 GATAACAAGCCTACAGCTGTTTTCAGACAGAGATTGCAATGCA 417
Db 118 GCAACAGATCCGATTCAGAGTCAGATCCAGAAAGTTGAGATGCA 74

RESULT 13
CR442585/c
LOCUS
DEFINITION
Xenopus tropicalis XGC-tailbud cDNA clone TTBa073h24 3',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus tropicalis (western clawed frog)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
Unpublished (2004)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TTBa073h24.q1k7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: T7.

FEATURES
source
Location/Qualifiers
1..926
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTBa073h24"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud"
/notes="Vector: pCSI07; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5' of poly A+ RNA from tailbud.
EcoRI-NotI cut cDNA was then ligated into pCSI07 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match 5.2%; Score 39; DB 7; Length 926;
Best Local Similarity 50.8%; Pred. No. 4.7; Mismatches 0; Indels 0; Gaps 0;
Matches 93; Conservative 0

QY 8 AGATGACCGGAATGGTAAACGGAAGAAGGTCTGTGCAATTTTATTTGCTGAGGCATTAAACAG 67
Db 186 AGAAGAACAGCAGCTTATAGATTCATGACATGATACCAAGGCTGCCTGAGTGCTA 127

QY 68 AAAACTGCAGAGAAATATTGAAACGCGCAGGAGCATGTTTGGTGGGATCAGCCCAATTA 127
Db 126 AGCTCTGAAGTCCAAACATTGTCTTGGATGATGTTTGTGGGTGGGGGTATATA 67

QY 128 ACAGCAGTTTTCAGAGGATATATTACAGATTAATTCATGATGGCGAAGCTCAATTTA 187
Db 66 GGGAGGGTTTTCAGAGGATTAATCAATNAATAAAGTAAATGTCAAAAAAATAAAAAA 7

QY 188 AAA 190
Db 6 AAA 4

RESULT 14
CNS017KX/c
LOCUS
DEFINITION
Xenopus tropicalis melanogaster genome survey sequence SP6 end of BAC
BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.

FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37F10"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : SP6"

ORIGIN
Query Match 5.2%; Score 39; DB 9; Length 1101;
Best Local Similarity 21.9%; Pred. No. 4.9; Mismatches 86; Conservative 135; Mismatches 165; Indels 7; Gaps 1;
Matches 134

QY 134 GTTTTCAGAGGATATATTACAGATTAATTCGATGGCGCAAGCTCAATTTAAAGCG 193
Db 1090 DRTWTKRKKKTKTKTADKKTRWDTKGGGGGKGGKGGKADTKRWRGAATR 1031

QY 194 TTTTCAGTTTACTTCGAGGCATGAGCGGCTAATCTCTAGAGCGCTTGGAACTCGA 253
Db 1030 RARWATAATATATADTTWTKRTTGTCTGTCGACGATTAATCGGGCGCGCG 971

QY 254 GAGGAAAGGCTGAACGAAAGTAAGGAAGAGAGTATCACGAAACAGGAGATTTCGAGAA 313
Db 970 RWTWRWATRGDGRAGRAKRRDRRGRRRRRRTKRWWRRAAWTTATTTWWTGTTK 911

QY 314 GAGCCCTTGTGGCTCATGCGGGGATCCGAAAGCGATTTCATACATTTTCTGATTTATAG 373
Db 910 GA-----WWTKTRKKGGAATAADWGRTDDWRWDKRAAAAAGKGGKGAATWGAA 858

QY 374 ATAAACAAAGCTACAGCTGTTGAGACAGAGAGTTGAACATGATCTTTTTCAGCAGCTC 433
Db 857 AAAAAAWDTKTATKTKRTKTGRAWAAAGTTRTTWDAATAWTKRAAAAAAGRAAAAWAK 798

QY 434 ATTTTCGACATGCTGTTTGGACATGTCGTCGACGATTAATCGGGCGCGCGCGCGCG 493
Db 797 DRGTRKGGKRGKRGTDGDKGTWKTKRTTWTKTAAARRAARAAAGKGTCTKTGRRARA 738

QY 494 TCAGTTTGTATGTAAGAAAGTTCAGTTCAGGATA 526
Db 737 WDAWTTTTRKGAATRDGAAAAAATTTGGGAAA 705

RESULT 15
CNS00DLS
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR27B02 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL072703

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:45:13 ; Search time 84.1768 Seconds
(without alignments)
1069.666 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNEMTGMVTRSRVHPIA.....YLIVQEMAQLSEKRITSEG 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_238ep04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1289	100.0	251	8	ADJ62045
2	400.5	31.1	252	6	ABM70364
3	249	19.3	239	8	ADH02827
4	103.5	8.0	645	2	AAW29653
5	102.5	8.0	557	4	AAU20133
6	102.5	8.0	557	5	ABG91382
7	102.5	8.0	616	6	ABR41881
8	102.5	8.0	616	6	AAE36781
9	102.5	8.0	645	6	ABR41872
10	102.5	8.0	645	6	ABR41875
11	102.5	8.0	645	6	AAE36782
12	102.5	8.0	645	7	ADG31694
13	102.5	8.0	645	7	ADG31692
14	97	7.5	1554	7	ABO78068
15	95.5	7.4	2329	2	AAW25038
16	95.5	7.4	2637	4	AAU31618
17	95.5	7.4	3485	4	AAU31615
18	94.5	7.3	2986	4	AAU31620
19	94.5	7.3	3418	2	AAW23287
20	94.5	7.3	3418	2	AAW19211
21	94.5	7.3	3418	2	AAW04357
22	94.5	7.3	3418	2	AAW04356
23	94.5	7.3	3418	2	AAW04358
24	94.5	7.3	3418	2	AAW04354
25	94.5	7.3	3418	2	AAW04355

26	94.5	7.3	3418	3	AAW77819	AAW77819 BRCA2 PRO
27	94.5	7.3	3418	8	ADJ32561	Adj32561 Human BRA
28	94.5	7.3	3418	8	ADK67819	Adk67819 Human BRC
29	94.5	7.3	3418	8	ADL32565	Adl32565 Human BRC
30	94.5	7.3	3423	4	ABG23417	ABG23417 Novel hum
31	94.5	7.3	4741	4	AAU31612	AAU31612 Novel hum
32	94.5	7.3	4743	4	AAU31619	AAU31619 Novel hum
33	94.5	7.3	5292	4	AAU31617	AAU31617 Novel hum
34	93.5	7.3	3418	7	ADJ68372	Adj68372 Human hea
35	93	7.2	480	4	ABB64489	Abb64489 Drosophila
36	90	7.0	198	5	ABB98195	Abb98195 Arabidops
37	90	7.0	198	5	AAW77963	AAW77963 Arabidops
38	90	7.0	198	6	ABG75910	ABG75910 Arabidops
39	90	7.0	198	7	ADP29071	Adp29071 Arabidops
40	90	7.0	200	5	AAU93026	AAU93026 Arabidops
41	90	7.0	200	7	ADC46691	Adc46691 Thalecres
42	90	7.0	200	7	ADD30082	Add30082 Plant yie
43	90	7.0	200	8	ADI41695	Adi41695 Plant tra
44	90	7.0	200	8	ADO02113	Ado02113 Thalecres
45	90	7.0	534	2	AAW25031	AAW25031 Partial B

ALIGNMENTS

RESULT 1
ADJ62045
ID ADJ62045 standard; protein; 251 AA.

XX AC ADJ62045;
XX DT 06-MAY-2004 (first entry)
XX DE Bacillus subtilis yvmC protein SEQ ID NO:8.
XX KW heterologous biological substance; Bacillus; red pigment.
XX OS Bacillus subtilis.
XX PN WO2004011609-A2.
XX PD 05-FEB-2004.
XX PF 25-JUL-2003; 2003WO-US023398.
XX PR 26-JUL-2002; 2002US-0398853P.
XX PA (NOVO) NOVOZYMES BIOTECH INC.
XX Tang M, Sloma A, Sternberg D, Behr R;
XX WPI: 2004-143839/14.
XX N-PSDB; ADJ62044.

Producing a heterologous biological substance comprises cultivating pigment-deficient mutants of Bacillus cell that directs synthesis of the heterologous biological substance and has a modification of the cypx and yvmC genes.

Example 1; SEQ ID NO'8; 62pp; English.

The present invention describes a method for producing a heterologous biological substance comprising cultivating a mutant of a parent Bacillus cell in a medium suitable for the production of a heterologous biological substance, and recovering the heterologous biological substance from the cultivation medium, where the mutant cell comprises a first nucleic acid sequence directing synthesis of the heterologous biological substance and a second nucleic acid sequence comprising a modification of at least one of the genes cypx and yvmC, which are involved in the production of a red pigment. The mutant cell is also deficient in the production of the red pigment compared to the parent Bacillus cell when cultivated under the same conditions. Also described: (1) a mutant of a parent Bacillus cell, comprising a first nucleic acid sequence directing synthesis of a

Handwritten signature

CC heterologous biological substance and a second nucleic acid sequence
CC comprising a modification of at least one of the genes cypX and ymcC,
CC which are involved in the production of a red pigment, wherein the mutant
CC cell is deficient in the production of the red pigment compared to the
CC parent Bacillus cell when cultivated under the same conditions; and (2)
CC obtaining a mutant of a parent Bacillus cell, comprising introducing into
CC the parent Bacillus cell a first nucleic acid sequence directing
CC synthesis of a heterologous biological substance and a second nucleic
CC acid sequence comprising a modification of at least one of the genes cypX
CC and ymcC, which are involved in the production of a red pigment, and
CC identifying the mutant cell comprising the modified nucleic acid
CC sequence, where the mutant cell is deficient in the production of the red
CC pigment compared to the parent Bacillus cell when cultivated under the
CC same conditions. The methods and compositions of the present invention
CC are useful for producing biological substances, e.g. nucleic acid,
CC polyamide, polyamine, polyol, polypeptide, or polysaccharide, in pigment-
CC deficient mutants of Bacillus cells, and obtaining the pigment-deficient
CC mutant cells. The present sequence represents Bacillus subtilis ymcC,
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 251 AA;

Query Match 100.0%; Score 1289; DB 8; Length 251;
Best Local Similarity 100.0%; Pred. No. 4.5e-143;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNEMTGMVTRRSVHFIAEALTENCREFERRRHVLVGISPFNSRSEDYIYRLIGWAKA 60
DB 1 MNEMTGMVTRRSVHFIAEALTENCREFERRRHVLVGISPFNSRSEDYIYRLIGWAKA 60
QY 61 QFKSVSVLLAGHEAANLLEALGTPRGKAERKVKESRNRFAERLVAHGGDPKAIHTF 120
DB 61 QFKSVSVLLAGHEAANLLEALGTPRGKAERKVKESRNRFAERLVAHGGDPKAIHTF 120
QY 121 SDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREAIIGRAGVSLMMEEVSEDMNLNLA 180
DB 121 SDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREAIIGRAGVSLMMEEVSEDMNLNLA 180
QY 181 VEVVIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQ 240
DB 181 VEVVIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEMAQ 240
QY 241 MLSEKRITSEG 251
DB 241 MLSEKRITSEG 251

RESULT 2
ABM70364
ID ABM70364 standard; protein; 252 AA.
XX
AC ABM70364;
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #3461.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
FN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.

(CNRS) CNRS CENT NAT RECH SCI.
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunet F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 3461; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 252 AA;
Query Match 31.1%; Score 400.5; DB 6; Length 252;
Best Local Similarity 37.6%; Pred. No. 3.5e-38;
Matches 86; Conservative 43; Mismatches 97; Indels 3; Gaps 2;
QY 13 SVHFIABALTEACREIFERRRHVLVGISPFNSRSEDYIYRLIGWAKAOKPSVLL-AG 71
DB 23 SPSFTVOGETSRCDQIIQGDHALIGISPFNSRSEDYIYRLIGWAKAOKPSVLL-AG 82
QY 72 HEANLLEALGTPRGKAERKVKESRNRFAERLVAHGGDPKAIHT--FSDFINKAY 129
DB 83 REASRLLVASGIDNVKAIKTHREIRHLRLNLDYVISTATLKSQKIRVIOFSDFSLNHDY 142
QY 130 QLILRQVEHAFPEQPHFRHACLDMSREAIIGRAGVSLMMEEVSEDMNLNLAIVYIAELP 189
DB 143 QSLKTOVENAFNESESFKKSLCDMSFOAIKGRKLGITGOYFGQIDQLVYKALPYIFAEIP 202
QY 190 FFTGAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLIVQEM 238
DB 203 FYLNTPLRLGVKYSTLLYHRPWSIGKFLNGSPYIOVADKQSYGIVTQL 251
RESULT 3
ADH02827
ID ADH02827 standard; protein; 239 AA.
XX
AC ADH02827;
XX
DT 11-MAR-2004 (first entry)
XX
DE S. noursei AlbC SEQ ID NO:9.
XX
KW AlbC; diketopiperazine; antibacterial; fungicide; virucide;
KW immunosuppressant; cytostatic; cyclodipeptide; CDP; alpha;
KW beta-unsaturated diketopiperazine; DKP; aliboursine.
XX
OS Streptomyces noursei.
XX

```
PN W02004000879-A1.
XX 31-DEC-2003.
XX 18-JUN-2003; 2003WO-FR001851.
XX 21-JUN-2002; 2002FR-00007728.
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX (CNRS ) CENT NAT RECH SCI.
XX Gondry M, Genet R, Lautru S, Pernodet J;
XX WPI; 2004-156410/15.
XX N-PSDB; ADH02821.
XX New nucleic acid encoding proteins involved in synthesis of
XX cycloolipeptides and diketopiperazine derivatives, useful e.g. as
XX antibacterial agents and preparation of encoded polypeptides.
XX Claim 22; SEQ ID NO 9; 53pp; French.
XX The invention relates to a novel isolated, natural or synthetic
XX polynucleotide and encoded polypeptide involved in the synthesis of
XX diketopiperazine derivatives. a protein of the invention has
XX antibacterial, fungicide, virucide, immunosuppressant, and cytostatic
XX activity. The polynucleotide and related sequences are useful as probes
XX and amplification primers, also for preparation of the encoded
XX polypeptides (II) and of modified biological and in vitro systems
XX systems. The polypeptides are useful for preparation of cyclodipeptides
XX (CDP) or alpha,beta-unsaturated diketopiperazine (DKP) derivatives
XX substituted at positions 3 and 6 by amino acid sidechains, most
XX particularly albonoursin. The present sequence is used in the
XX exemplification of the invention.
XX Sequence 239 AA;
Query Match 19.3%; Score 249; DB 8; Length 239;
Best Local Similarity 29.6%; Pred. No. 2.5e-20;
Matches 67; Conservative 39; Mismatches 110; Indels 10; Gaps 4;
QY 19 EALTENCREFERRHVLVGISPNRFSFSDIYRLIGWAKAPKSVSVLLAGHEAANLL 78
DB 16 EILGDRSLIRQGEHALIGISAGNSYFSQKNTVMLLQWAGQFERTDVVVVDHDEML 75
QY 79 EALGTPRGKAERKVRKVSRRRFAERALVAHGDPK--AIHTPSDFIDNKAYQLLRQEV 136
DB 76 IADGRSAQEAERSVKRTLKDLRRRLRSLSGVGDHAERFVRSLSELQETPEYRAVRERT 135
QY 137 EHAFPEQPHFRHACLDMSREAIIGR-ARGVSLMMEEVSEDMNLAVYVIAELPFFI 195
DB 136 DRAFEEDAEEFATACEDMVRVAVVNRPGDVG-ISAELHLAGLVYLAELPLFADSP 190
QY 196 DILEVEETLAYHRPWKLGEKISNHEFSICMRPN--QGYLIVQEMAOQLSEKRITSEG 231
DB 191 GVFSVPSSVLCYHIDTPTITAFLSRRRTGFRAAEQAVVVRPQSLA 236
RESULT 4
AAW29653
ID AAW29653 standard; protein; 645 AA.
XX AC AAW29653;
XX 09-NOV-1998 (first entry)
XX Human secreted protein DF518_3.
XX Secreted protein; DF518_3; human.
XX Homo sapiens.
XX W09830582-A2.
XX PN
16-JUL-1998.
XX PD XX
XX PF 09-JAN-1998; 98WO-US000289.
XX PR 09-JAN-1997; 97US-00780890.
XX PR 08-JAN-1998; 98US-00004680.
XX (GEMY ) GENETICS INST INC.
XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
XX Spaulding V, Agostino MJ;
XX WPI; 1998-413681/35.
XX N-PSDB; AAV40506.
XX New isolated nucleic acids and secreted proteins - obtained from human
XX foetal kidney, human adult retina, human foetal brain, human adult brain
XX and human adult blood cDNA libraries.
XX Claim 27; Page 74-77; 103pp; English.
XX This is the amino acid sequence of novel human secreted protein DF518_3
XX as predicted from a human adult brain cDNA clone (see AAV40506). The
XX clone was isolated using methods which are selective for cDNAs encoding
XX secreted proteins, or was identified as encoding a secreted or
XX transmembrane protein on the basis of computer analysis of the amino acid
XX sequence of the encoded protein. The invention relates to 9 cDNA clones
XX (see AAV40501-09), all deposited as ATCC 98290, that code for human
XX secreted proteins (see AAV29648-56) of the foetal kidney or brain, or
XX adult retina, brain or blood. Mammalian host cells and methods of
XX producing the (especially mature) polypeptides are claimed. The
XX polynucleotides and polypeptides can be used as e.g. nutritional sources
XX or supplements or may exhibit e.g. cytokine and cell proliferation or
XX differentiation activity, immunostimulant or immunosuppressive activity,
XX haematopoiesis regulating activity, receptor/ligand activity,
XX antiinflammatory activity, cadherin or tumour invasion suppressor
XX activity, tumour inhibition activity or other activities. DF518_3 protein
XX shows some sequence similarity to a number of database sequences
XX Sequence 645 AA;
Query Match 8.0%; Score 103.5; DB 2; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.016;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGKAEKVRKEVSRNRRFAERALVAHGDPKAIHTPSDFIDNKAYQLLRQEVHAFEEQ 143
DB 183 PEGAEKTKVKIEKLAR-----FVAGGPELEKVMEDYKDNPAFLHDKNSRGFL-- 232
QY 144 PFRHACLDMSREAIIGRARGVSLMMEEVSEDMNLAVYVIAELPFFI--GAPDILEVE 201
DB 233 -YRKVKAERKQAQKSAQKVSPPF-DEEVKNLA-----EKLARFIADGGP--EVE 282
QY 202 ETLLAVHRPWKLGEKISNHEFSICMRPN--QGYLIVQEMAOQLSEKRITSEG 251
DB 283 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 325
RESULT 5
AAU20133
ID AAU20133 standard; protein; 557 AA.
XX AC AAU20133;
XX 06-DEC-2001 (first entry)
XX Human DNA repair and processing polypeptide #18.
XX DNA processing; human; mouse; rabbit; goat; horse; cat; gene therapy;
XX chicken; sheep; immunosuppressive; antithratic; vasotropic;
XX antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
XX cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
XX KW
```


08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254037P.
03-JAN-2001; 2001US-02559678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-496846/54.
N-PSDB; AAS32200.
Nucleic acids encoding 41 human polypeptides, useful for preventing,
diagnosing and/or treating e.g. cancers, Parkinson's disease and diabetic
retinopathy.
Claim 11; SEQ ID NO 60; 460pp; English.
Sequences AAU20116-AAU20147 represent the DNA repair and processing
polypeptides of the invention. DNA repair and processing polypeptides and
their associated polynucleotides are useful in the diagnosis, treatment
and prevention of various types of disorders in e.g. humans, mice,
rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological
condition can be determined by detecting the presence or absence of a
mutation in a DNA repair and processing polynucleotide. The treatable
disorders include autoimmune diseases such as rheumatoid arthritis,
hyperproliferative disorders such as neoplasms of the breast or liver,
cardiovascular disorders such as cardiac arrest, cerebrovascular
disorders such as cerebral ischaemia, nervous system disorders such as
Alzheimer's disease, infections caused by bacteria, viruses and fungi,
ocular disorders such as corneal infection, endocrine disorders such as
premature labour and infertility, gastrointestinal disorders such as
Crohn's disease, renal disorders such as glomerulonephritis and
respiratory disorders such as asthma and pleurisy. The polypeptides can
also be used to aid wound healing, to prevent skin aging due to sunburn,
to maintain organs before transplantation, to regenerate tissues and in
chemotaxis. Note: The sequence data for this patent did not form part of
the printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 8.0%; Score 102.5; DB 4; Length 557;
Best Local Similarity 26.3%; Pred. No. 0.017;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKVRNRRFAERLVAHGDPKAIHTFSDFDIDNKAYQLLRQEVHAFPEQ 143
Db 95 PEGAEIRKVRKVRNRRFAERLVAHGDPKAIHTFSDFDIDNKAYQLLRQEVHAFPEQ 144
QY 144 PHERHACLDMSREAIIGRAGVSLMMEVESEDMNLAVEYVIAELPFI--GAPDILEVE 201
Db 145 -YRKKVAIRKEAKQSAQSKVSPPE-DEEVKNLA-----EKLAPIDAGGP---EVE 194
QY 202 ETLLAYHPKWLGEKINSNHFISICWRPN-QCYLLVQEMQMLSEKRTISEG 251
Db 195 TIALQNNR-----ENQAFSLYEYNSQGYKYRQKLEFRKAKASSTG 237

RESULT 6
ABG91382
ID ABG91382 standard; protein; 557 AA.
AC ABG91382;
XX
DT 29-NOV-2002 (first entry)
XX
DE Novel human DNA repair protein #18.
XX Human; DNA repair; immunodeficiency; autoimmune disorder; diabetes;
KW rheumatoid arthritis; multiple sclerosis; allergy; asthma; cancer;
KW inflammatory condition; blood-related disorder; renal disorder;

KW hyperproliferative disorder; cardiovascular disorder; Crohn's disease;
KW respiratory disorder; neurological disorder; endocrine disorder;
KW reproductive system disorder; infectious disease; neuronal damage;
KW gastrointestinal disorder; neuro-degenerative; keratinocyte growth;
KW hair loss; nutritional; gene therapy; forensic; immunophenotype.
XX Homo sapiens.
OS
XX
PN US2002086353-A1.
XX
PD 04-JUL-2002.
XX
PF 17-JAN-2001; 2001US-00764856.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216847P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229387P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-023935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX WPI; 2002-642242/69.
DR N-PSDB; ABS67503.
DR

```

XX Novel polypeptide useful for diagnosis, prognosis, prevention, and
PT treatment of immune, hyperproliferative, renal, respiratory,
PT cardiovascular, reproductive, endocrine, gastrointestinal and
PT neurological disorders.
XX
PS Claim 11; Page 207-209; 296pp; English.
PS
XX The invention relates to novel isolated polypeptides (I) and
XX polynucleotides (II). (I) and (II) and antibody (III) to (I) are useful
CC for diagnosing a pathological condition or susceptibility to a
CC pathological condition in a subject and for preventing, treating or
CC ameliorating a medical condition. (I) is also useful for identifying a
CC binding partner to the polypeptide. (I), (II) and (III) are useful in
CC treating, preventing, diagnosing and/or prognosing immunodeficiencies
CC (e.g., B cell immunodeficiencies, severe combined immunodeficiencies),
CC autoimmune disorders (rheumatoid arthritis, multiple sclerosis, diabetes
CC mellitus), allergic reactions and conditions (e.g., asthma), inflammatory
CC conditions, graft-versus-host disease, blood-related disorders
CC (thrombosis, atherosclerosis), hyperproliferative disorders (e.g.,
CC cancer), renal disorders (e.g., acute glomerulonephritis), cardiovascular
CC disorders (e.g., arrhythmia), respiratory disorders (Goodpasture's
CC syndrome), neurological disorders (e.g., Alzheimer's disease, Parkinson's
CC disease), endocrine disorders (e.g., Addison's disease), reproductive
CC system disorders (e.g., endometriosis), infectious diseases (e.g., viral,
CC bacterial or fungal infections), and gastrointestinal disorders (e.g.,
CC Crohn's disease) (I) is useful to stimulate neuronal growth and treat,
CC prevent, and/or diagnose neuronal damage which occurs in certain neuronal
CC disorders or neuro-degenerative conditions, for stimulating keratinocyte
CC growth, to prevent hair loss, to modulate mammalian characteristics such
CC as body height, weight, hair colour, and to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors or other nutritional components. (I) is also useful
CC as a protein molecular weight marker and to raise antibodies. (II) is
CC useful for chromosome identification, radiation hybrid mapping, in gene
CC therapy, for identifying individuals from minute biological samples, as
CC additional DNA markers for restriction fragment length polymorphism
CC (RFLP), in forensic biology, as molecular weight markers on Southern
CC gels, as diagnostic probes for the presence of a specific mRNA in a
CC particular cell type, to raise anti-DNA antibodies using DNA immunisation
CC techniques, and as an antigen to elicit an immune response. (III) is
CC useful for immunophenotyping cell lines and biological samples and for
CC diagnosing and treating diseases, disorders or conditions. (III) is also
CC useful to assay protein levels in a biological sample. ABG91364-ABG91396
CC represent human amino acid sequences of the invention
XX
SQ Sequence 557 AA;
Query Match 8.0%; Score 102.5; DB 5; Length 557;
Best Local Similarity 26.3%; Pred. No. 0.017;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKEVSRNRRFAERALVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFPEQ 143
Db 95 PEGATRKVIEKLAR-----FVAGGPELEKVMEDYKDNPAFLDKNSREFL-- 144
QY 144 PFRHACLMSREAIIGRAGVSLMMEVESEDMNLNLAVEVYIAELPPFI--GAPDILEVE 201
Db 145 -YRKVKVAIRKEAQKSAQASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 194
QY 202 ETLLAYHRPWKLGKESINHEFSICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 195 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEFRKAKASSTG 237
RESULT 7
ABR41881
ID ABR41881 standard; protein; 616 AA.
XX
AC ABR41881;
XX
DT 11-AUG-2003 (first entry)
XX
DE Human caspase-8 binding protein (p72) EST clone, THCS10568.

```

```

DE Predicted protein sequence of a human EST clone.
XX Human; Cari; caspase-8 interacting protein; apoptosis; neuroprotective;
XX immunosuppressive; antidiabetic; hepatotropic; virucide;
XX antiinflammatory; antitumor; cytostatic; gene therapy;
XX expressed sequence tag; EST.
XX
OS Homo sapiens.
XX
XX WO2003020759-A2.
XX
XX 13-MAR-2003.
XX
XX 04-SEP-2002; 2002WO-IL000733.
XX
XX 04-SEP-2001; 2001IL-00145278.
XX
XX 31-OCT-2001; 2001IL-00146251.
XX
XX 06-JAN-2002; 2002IL-00147487.
XX
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Wallach D, Goncharov T, Kolumam G, Rajput A;
XX WPI; 2003-300863/29.
XX
XX Novel intracellular caspase-8 interacting polypeptide, designated as Cari
XX polypeptide, useful for treating inflammatory disease including multiple
XX sclerosis, autoimmune uveoretinitis and diabetes.
XX
XX Example 7; Page 116-119; 124pp; English.
XX
XX The present sequence is the predicted protein sequence of a human
XX expressed sequence tag (EST) clone obtained from a database screening
XX using tryptic peptides of an isolated 72.5 kDa protein identified as
XX Cari. The EST was used to screen a human gene database, and a cDNA (see
XX ACC48539) encoding Cari (see ABR41872) was subsequently obtained. Cari is
XX a novel intracellular caspase-8 interacting polypeptide. Cari
XX polypeptides, fragments, mutants, variants and derivatives, nucleic acids
XX encoding them, and anti-Cari antibodies can be used to treat an
XX inflammatory disease such as multiple sclerosis with primary
XX oligodendroglipathy, autoimmune uveoretinitis, diabetes, lupus,
XX autoimmune myocarditis I, hepatitis C virus (HCV) mediated chronic
XX hepatitis, chronic gastritis e.g. type A gastritis, mixed connective
XX tissue disease (MCTD), Crohn's disease and ulcerative colitis
XX
XX Sequence 616 AA;
Query Match 8.0%; Score 102.5; DB 6; Length 616;
Best Local Similarity 26.3%; Pred. No. 0.02;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKEVSRNRRFAERALVAHGDPKAIHTFSDFDINKAYQLLRQVEVHAFPEQ 143
Db 154 PEGATRKVIEKLAR-----FVAGGPELEKVMEDYKDNPAFLDKNSREFL-- 203
QY 144 PFRHACLMSREAIIGRAGVSLMMEVESEDMNLNLAVEVYIAELPPFI--GAPDILEVE 201
Db 204 -YRKVKVAIRKEAQKSAQASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 253
QY 202 ETLLAYHRPWKLGKESINHEFSICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 254 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEFRKAKASSTG 296
RESULT 8
AAE36781
ID AAE36781 standard; protein; 616 AA.
XX
AC AAE36781;
XX
DT 07-AUG-2003 (first entry)
XX
DE Human caspase-8 binding protein (p72) EST clone, THCS10568.

```

XX Human; immunisation; caspase-8; enzyme linked immunosorbent assay; ELISA;
KW immunofluorescence; immunoelectron microscopy; caspase-8 binding protein;
KW p72 protein.
XX Homo sapiens.
XX WO2003020767-A2.
XX PD 13-MAR-2003.
XX 04-SEP-2002; 2002WO-IL000734.
XX PF 04-SEP-2001; 2001IL-00145279.
XX PR (YEDA) YEDA RES & DEV CO LTD.
XX PA Wallach D, Goncharov T, Kolumam G;
XX WI; 2003-290170/28.
XX Novel antibody which co-immunoprecipitates caspase together with caspase-
PT bound protein, and releases caspase and bound protein from immune complex
PT upon elution, useful for purifying caspase-8 or caspase-bound protein.
XX Example 6; Page 68-70; 75pp; English.
XX The present invention relates to antibodies obtainable by immunisation of
CC an animal with a peptide from the C terminus end of the caspase-8 sub-1
CC unit and its fragments, capable of co-immunoprecipitating caspase (both
CC active caspase-8 and pro-caspase-8) together with a caspase-bound protein
CC and of releasing the caspase and the bound protein efficiently from the
CC immune complex upon elution. The antibodies of the invention are useful
CC for developing complex of an enzyme linked immunosorbent assay (ELISA) and for
CC purifying caspase (preferably caspase-8) or caspase-bound protein. They
CC are useful to effectively co-precipitate and isolate a protein bound to
CC caspase-8. The antibodies can be employed histologically as in immuno-
CC fluorescence or immunoelectron microscopy for the in situ detection of
CC caspase-8. The present sequence is human caspase-8 binding protein (p72)
CC EST clone, THCS10568. This sequence is used in the exemplification of the
CC invention
XX
SQ Sequence 616 AA;
Query Match 8.0%; Score 102.5; DB 6; Length 616;
Best Local Similarity 26.3%; Pred. No. 0.02;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKEVSRNRRFAERLVAHGDPKAIHTFSDIDNKAYQLLQVEHAFPEQ 143
Db 154 PEGAEIRKVKIEKLAR-|-----FVARGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 203
QY 144 PHERHACLDMSREAIIGRARGVSLMMEVESEDMNLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 204 -YRKKVAEIRKKAQSKQAASQKVSPPPE-DEEVKNLA-----EKLFARFIADGGP---EVE 253
QY 202 ETLLAYHRPWKLGKESINHEFSICMRPN-QGYLIVQEMQMLSEKRTITSEG 251
Db 254 TIALQNNR-----ENQAFSFLYEPNQSGYKYRQKLEEFKAKASSTG 296
RESULT 9
ID ABR41872
XX ABR41872 standard; protein; 645 AA.
AC ABR41872;
XX 11-AUG-2003 (first entry)
XX Human caspase-8 interacting protein, Cari.
XX Human; Cari; caspase-8 interacting protein; apoptosis; neuroprotective;
KW immunosuppressive; antidiabetic; hepatotropic; virucide;

XX antiinflammatory; antiulcer; cytostatic; gene therapy.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Domain 414..437
FT /note= "binds caspase-8, given in ABR41873"
FT Domain 422..437
FT /note= "binds caspase-8, given in ABR41874"
XX WO2003020759-A2.
XX PN 13-MAR-2003.
XX 04-SEP-2002; 2002WO-IL000733.
XX PF 04-SEP-2001; 2001IL-00145278.
XX PR 31-OCT-2001; 2001IL-00146251.
XX PR 06-JAN-2002; 2002IL-00147487.
XX PA (YEDA) YEDA RES & DEV CO LTD.
XX Wallach D, Goncharov T, Kolumam G, Rajput A;
XX WI; 2003-300863/29.
XX N-PSDB; ACC48539.
XX Novel intracellular caspase-8 interacting polypeptide, designated as Cari
PT polypeptide, useful for treating inflammatory disease including multiple
PT sclerosis, autoimmune uveoretinitis and diabetes.
XX Claim 1; Page 120-122; 124pp; English.
XX The present sequence is the protein sequence of human Cari, a novel
CC intracellular caspase-8 interacting polypeptide. Cari activity is
CC required for induction of apoptosis through the tumour necrosis factor
CC receptor signalling pathway or by caspase-8. It binds to pro-caspase8 but
CC not to active caspase-8. Claimed pharmaceutical compositions comprising
CC Cari, its mutein, isoform, allelic variant, fragment, fusion protein or
CC derivative, a DNA sequence encoding these, an antisense or ribozyme
CC sequence, a vector comprising the DNA, or an anti-Cari antibody, are used
CC in the treatment of an inflammatory disease such as multiple sclerosis
CC with primary oligodendroglialopathy, autoimmune uveoretinitis, diabetes,
CC lupus, autoimmune myocarditis I, hepatitis C virus (HCV) mediated chronic
CC hepatitis, chronic gastritis e.g. Type A gastritis, mixed connective
CC tissue disease (MCTD), Crohn's disease and ulcerative colitis. A claimed
CC method for isolating polypeptides or factors involved in intracellular
CC signalling processes uses a Cari polypeptide, mutein, isoform, allelic
CC variant, fragment, fusion protein or derivative as the prey or bait in a
CC yeast two-hybrid procedure. A claimed method for screening for a peptide
CC or a small molecule antagonist to Cari comprises high-throughput
CC screening and selection of such molecules able to inhibit the interaction
CC of Cari to pro-caspase-8, or able to inhibit apoptosis enhanced by Cari
XX
SQ Sequence 645 AA;
Query Match 8.0%; Score 102.5; DB 6; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKEVSRNRRFAERLVAHGDPKAIHTFSDIDNKAYQLLQVEHAFPEQ 143
Db 183 PEGAEIRKVKIEKLAR-|-----FVARGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 232
QY 144 PHERHACLDMSREAIIGRARGVSLMMEVESEDMNLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 233 -YRKKVAEIRKKAQSKQAASQKVSPPPE-DEEVKNLA-----EKLFARFIADGGP---EVE 282
QY 202 ETLLAYHRPWKLGKESINHEFSICMRPN-QGYLIVQEMQMLSEKRTITSEG 251
Db 283 TIALQNNR-----ENQAFSFLYEPNQSGYKYRQKLEEFKAKASSTG 325

```
RESULT 10
ABR41875
ID ABR41875 standard; protein; 645 AA.
AC
AC ABR41875;
DT 11-AUG-2003 (first entry)
DE Human caspase-8 interacting protein, Cari, D600E mutin.
DE
XX
XX Human; Cari; caspase-8 interacting protein; apoptosis; neuroprotective;
KW immunosuppressive; antidiabetic; hepatotropic; virucide;
KW antiinflammatory; antiulcer; cytostatic; gene therapy; mutant; mutin.
XX
XX Homo sapiens.
OS
OS Synthetic.
FH
FH Key Location/Qualifiers
FT Misc-difference 600
FT /note= "wild-type Asp replaced by Glu"
XX
XX WO2003020759-A2.
PN
PN 13-MAR-2003.
PD
PD 04-SEP-2002; 2002WO-IL000733.
PF
PF 04-SEP-2001; 2001IL-00145278.
PR
PR 31-OCT-2001; 2001IL-00146251.
PR
PR 06-JAN-2002; 2002IL-00147487.
XX
XX (YEDA ) YEDA RES & DEV CO LTD.
PA
XX Wallach D, Goncharov T, Kolumam G, Rajput A;
XX WPI; 2003-300863/29.
XX
XX Novel intracellular caspase-8 interacting polypeptide, designated as Cari
PT polypeptide, useful for treating inflammatory disease including multiple
PT sclerosis, autoimmune uveoretinitis and diabetes.
XX
XX Claim 12; Page; 124pp; English.
XX
XX The present sequence is the protein sequence of a mutin of human Cari in
CC which the native Asp residue at position 600 is replaced by Glu. Cari is
CC a novel intracellular caspase-8 interacting polypeptide. Its activity is
CC required for induction of apoptosis through the tumour necrosis factor
CC receptor signalling pathway or by caspase-8. The D600E mutin is not
CC cleaved by caspase-8, giving it a dominant negative effect on the
CC activity of endogenous Cari, and increasing its cytotoxicity relative to
CC the native protein. Claimed pharmaceutical compositions comprising Cari,
CC its mutin (especially the D600E mutin), isoform, allelic variant,
CC fragment, fusion protein or derivative, a DNA sequence encoding these, an
CC antisense or ribozyme sequence, a vector comprising the DNA, or an anti-
CC Cari antibody, are used in the treatment of an inflammatory disease
CC selected from multiple sclerosis with primary oligodendroglialopathy,
CC autoimmune uveoretinitis, diabetes, lupus, autoimmune myocarditis I,
CC hepatitis C virus (HCV) mediated chronic hepatitis, chronic gastritis
CC e.g. type A gastritis, mixed connective tissue disease (MCTD), Crohn's
CC disease and ulcerative colitis. A claimed method for isolating
CC polypeptides or factors involved in intracellular signalling processes
CC uses a Cari polypeptide, mutin, isoform, allelic variant, fragment,
CC fusion protein or derivative as the prey or bait in a yeast two-hybrid
CC procedure. A claimed method for screening for a peptide or a small
CC molecule antagonist to Cari comprises high-throughput screening and
CC selection of such molecules able to inhibit the interaction of Cari to
CC pro-caspase-8, or able to inhibit apoptosis enhanced by Cari. Note: The
CC present sequence is not shown in the specification but is derived from
CC the human Cari sequence given in the sequence listing (see ABR41872)
XX
XX Sequence 645 AA;
XX Query Match 8.0%; Score 102.5; DB 6; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
```

```
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGKAERKVRKESVNRRAERALVAHGDDPKAIHTFSDFDIDNKAYQLLRQVEHAFPEQ 143
DB 183 PEGAEATRKVIKRLAR-----FVAEGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 232
QY 144 PHFRHACLDMSREALIGRARGVSLMMEVEVSDMLNLAVEYVIAELPFFI--GAPDILEVE 201
DB 233 -YIRKVAIRKEAQKQSAOSKVSPPPE--DEEVNLA-----EKLARFIADGGP---EVE 282
QY 202 ETLLAYHRWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRITSG 251
DB 283 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEFRKAKASSTG 325

RESULT 11
AAE36782
ID AAE36782 standard; protein; 645 AA.
XX
AC AAE36782;
XX
DT 07-AUG-2003 (first entry)
XX
DE Human caspase-8 binding protein (p72) .
XX
XX Human; immunisation; caspase-8; enzyme linked immunosorbent assay; ELISA;
KW immunofluorescence; immunoelectron microscopy; caspase-8 binding protein;
KW p72 protein.
XX
XX Homo sapiens.
XX
XX WO2003020767-A2.
PN
XX 13-MAR-2003.
PD
PF 04-SEP-2002; 2002WO-IL000734.
XX
XX 04-SEP-2001; 2001IL-00145279.
PR
XX (YEDA ) YEDA RES & DEV CO LTD.
XX
XX Wallach D, Goncharov T, Kolumam G;
XX WPI; 2003-290170/28.
DR
DR N-PSDB; AAB55651.
XX
XX Novel antibody which co-immunoprecipitates caspase together with caspase-
PT bound protein, and releases caspase and bound protein from immune complex
PT upon elution, useful for purifying caspase-8 or caspase-bound protein.
XX
XX Example 8; Page 72-74; 75pp; English.
XX
XX The present invention relates to antibodies obtainable by immunisation of
CC an animal with a peptide from the C terminus end of the caspase-8 sub-1
CC unit and its fragments, capable of co-immunoprecipitating caspase (both
CC active caspase-8 and pro-caspase-8) together with a caspase-bound protein
CC and of releasing the caspase and the bound protein efficiently from the
CC immune complex upon elution. The antibodies of the invention are useful
CC for development of an enzyme linked immunosorbent assay (ELISA) and for
CC purifying caspase (preferably caspase-8) or caspase-bound protein. They
CC are useful to effectively co-precipitate and isolate a protein bound to
CC caspase-8. The antibodies can be employed histologically as in immuno-
CC fluorescence or immunoelectron microscopy for the in situ detection of
CC caspase-8. The present sequence is human caspase-8 binding protein (p72) .
CC This sequence is used in the exemplification of the invention
XX
XX Sequence 645 AA;
XX Query Match 8.0%; Score 102.5; DB 6; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
```

QY 84 PRGAERKVRKEVSRNRFAERALVAHGDPKAIHTFSDFDINKAYQLLRQVEHAFEEQ 143
Db 183 PEGAETRKVIEKLAR-----FVAGGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 232
QY 144 PHERHACLDMSREAIIGRAGVSLMMEVESEDMLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 233 -YRKKAIRKEAQAQSQAASQKVSPPPE-DEEVKNLA-----EKARFIADGGP---EVE 282
QY 202 ETLLAYHRPWKLGEKISNHEFSCMRPN-OGYLIVQEMAOQLSEKRTITSEG 251
Db 283 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEEFKAKASSTG 325
RESULT 12
ADG31694
ID ADG31694 standard; protein; 645 AA.
AC ADG31694;
XX
DT 26-FEB-2004 (first entry)
XX
DE Human protein with VEGF promoting activity SeqID40.
XX
KW human; vascularisation; vascular endothelial growth factor; VEGF;
KW tumour metastasis; psoriasis; chronic rheumatoid arthritis;
KW diabetic retinopathy; angiofibroma; macular degeneration;
KW angina pectoris; cerebral infarction; vascular dementia;
KW Burger's disease; antiinflammatory; cytostatic; antiarthritic;
KW antipsoriatic; antiarteriosclerotic; cardiant; vasotropic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
PN WO2003083117-A2.
XX
PD 09-OCT-2003.
XX
PF 01-APR-2003; 2003WO-JP004163.
XX
PR 01-APR-2002; 2002JP-00098679.
PR 02-APR-2002; 2002US-0368977P.
XX
XX (ASAH) ASAH KASEI KK.
XX
PI Nagumo H, Matsuda A, Sugahara T;
XX
DR WPI; 2003-812544/76.
DR N-PSDB; ADG31693.
XX
PT New purified protein, useful for preparing a composition for treating or
PT preventing a disease associated with abnormal expression of VEGF or VEGF
PT receptor, e.g., solid tumor, inflammation, psoriasis or chronic
PT rheumatoid arthritis.
XX
PS Claim 1; SEQ ID NO 40; 405pp; English.
XX
XX This invention relates to novel proteins having vascularisation activity.
CC Specifically, it refers to proteins having an activity of promoting the
CC expression of vascular endothelial growth factor (VEGF) and/ or the VEGF
CC receptor. VEGF is a representative growth factor involved in normal
CC vascularisation in growing tissues, for example for fetal growth, tissue
CC repair and for the menstrual cycle and pregnancy. The present invention
CC describes these purified proteins, vectors and appropriate antibodies, as
CC well as a screening assay for identifying substances capable of
CC inhibiting or promoting VEGF (and VEGF receptor) expression. Furthermore,
CC it refers to compositions useful for the diagnosis, treatment and
CC prevention of diseases associated with abnormal VEGF/ VEGF receptor
CC expression, such as tumour metastasis, psoriasis, chronic rheumatoid
CC arthritis, diabetic retinopathy, angiofibroma, macular degeneration,
CC angina pectoris, cerebral infarction, vascular dementia or Burger's
CC disease. Accordingly, they exhibit antiinflammatory, cytostatic,
CC antiarthritic, antipsoriatic, antiarteriosclerotic, cardiant, vasotropic
CC and neuroprotective activities. This polypeptide sequence is a human

CC protein with VEGF promoting activity of the invention.
XX
SQ Sequence 645 AA;
Query Match 8.0%; Score 102.5; DB 7; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKEVSRNRFAERALVAHGDPKAIHTFSDFDINKAYQLLRQVEHAFEEQ 143
Db 183 PEGAETRKVIEKLAR-----FVAGGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 232
QY 144 PHERHACLDMSREAIIGRAGVSLMMEVESEDMLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 233 -YRKKAIRKEAQAQSQAASQKVSPPPE-DEEVKNLA-----EKARFIADGGP---EVE 282
QY 202 ETLLAYHRPWKLGEKISNHEFSCMRPN-OGYLIVQEMAOQLSEKRTITSEG 251
Db 283 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEEFKAKASSTG 325
RESULT 13
ADG31692
ID ADG31692 standard; protein; 645 AA.
XX
AC ADG31692;
XX
DT 26-FEB-2004 (first entry)
XX
DE Human protein with VEGF promoting activity SeqID38.
XX
KW human; vascularisation; vascular endothelial growth factor; VEGF;
KW tumour metastasis; psoriasis; chronic rheumatoid arthritis;
KW diabetic retinopathy; angiofibroma; macular degeneration;
KW angina pectoris; cerebral infarction; vascular dementia;
KW Burger's disease; antiinflammatory; cytostatic; antiarthritic;
KW antipsoriatic; antiarteriosclerotic; cardiant; vasotropic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
PN WO2003083117-A2.
XX
PD 09-OCT-2003.
XX
PF 01-APR-2003; 2003WO-JP004163.
XX
PR 01-APR-2002; 2002JP-00098679.
PR 02-APR-2002; 2002US-0368977P.
XX
XX (ASAH) ASAH KASEI KK.
XX
PI Nagumo H, Matsuda A, Sugahara T;
XX
DR WPI; 2003-812544/76.
DR N-PSDB; ADG31691.
XX
PT New purified protein, useful for preparing a composition for treating or
PT preventing a disease associated with abnormal expression of VEGF or VEGF
PT receptor, e.g., solid tumor, inflammation, psoriasis or chronic
PT rheumatoid arthritis.
XX
PS Claim 1; SEQ ID NO 38; 405pp; English.
XX
XX This invention relates to novel proteins having vascularisation activity.
CC Specifically, it refers to proteins having an activity of promoting the
CC expression of vascular endothelial growth factor (VEGF) and/ or the VEGF
CC receptor. VEGF is a representative growth factor involved in normal
CC vascularisation in growing tissues, for example for fetal growth, tissue
CC repair and for the menstrual cycle and pregnancy. The present invention
CC describes these purified proteins, vectors and appropriate antibodies, as
CC well as a screening assay for identifying substances capable of
CC inhibiting or promoting VEGF (and VEGF receptor) expression. Furthermore,

CC it refers to compositions useful for the diagnosis, treatment and
CC prevention of diseases associated with abnormal VEGF/ VEGF receptor
CC expression, such as tumor metastasis, psoriasis, chronic rheumatoid
CC arthritis, diabetic retinopathy, angiofibroma, macular degeneration,
CC angina pectoris, cerebral infarction, vascular dementia or Buerger's
CC disease. Accordingly, they exhibit antiinflammatory, cytostatic,
CC antiarthritic, antipsoriatic, antiarteriosclerotic, cardiant, vasotropic
CC and neuroprotective activities. This polypeptide sequence is a human
CC protein with VEGF promoting activity of the invention.

XX Sequence 645 AA;

Query Match 8.0%; Score 102.5; DB 7; Length 645;
Best Local Similarity 26.3%; Pred. No. 0.022;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGAERKVRKESVRRRPAERALVAHGDPKAIHTFSDFDINKAYQLLQVEHAFPEQ 143
DB 183 PEGATRKVTEKLAR-----FVABGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 232
QY 144 PHFRHACLDMSRAIIGRARGVSLMMEVSEDMNLNLAVEVYIAELPFFI--GAPDILEVE 201
DB 233 -YRKVVAETIRKEAQKSAQASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 282
QY 202 ETLLAVHRPWKLGKISNHEFSICMRPN-OGYLIVQEMAQMLSEKRITSEG 251
DB 283 TIALQNNR-----ENQAFSLYEPNSQYKYRQKLEPRKAKASSTG 325

RESULT 14

ABO78068
ID ABO78068 standard; protein; 1554 AA.

XX ABO78068;

XX 29-JUL-2004 (first entry)

XX Pseudomonas aeruginosa polypeptide #10243.

XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

XX Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

XX 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI; 2003-615309/58.

XX N-PSDS; ABD11639.

XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 26814; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant

CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABO67826-
CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html

XX Sequence 1554 AA;

Query Match 7.5%; Score 97; DB 7; Length 1554;
Best Local Similarity 19.4%; Pred. No. 0.36;
Matches 64; Conservative 33; Mismatches 95; Indels 138; Gaps 14;
QY 1 MMEMTGVTTERSIVHIAEALTENCREIFERRRHV-----LVGISPNSRFSEDIYRLI 55
DB 336 LGDATDAATDHADLHFVVAQAERALQSLERTAHIGLEDDVQGLLLFLAHVLED-VFOLA 394
QY 56 GWAKAQFKSVSVLLAGH-EAANLLEALGTPRGAERKVRK-----EVSRRNR--FAER-- 105
DB 395 GMGTROFHTTELALTEQGFAGLL--LVGDHGHVLTGIRTVQAEGLDRDRRAGFLDLRLA 452
QY 106 ALVAHGDPKAIHT-----
DB 453 VLVEHGADTTVVDADQHHVALAQGTVDQHGHHRAASLVQARLDHHTATARGRSQFOFEH 512
QY 120 -----FSDFD-----NKAYQLLQVEHA-----PFEQPH 145
DB 513 FRLQQRFEQFTDASTELRRNRDEGSVAAPLFRDHVESQAVLDVVRVGLGLVDLVHRDH 572
QY 146 FRHA-----CLDMSREAIIGR-----ARGV-----SLMM 169
DB 573 DRHARRFVLHGLGHRDAVGRDHQHDHIDIGRLGATGTHRGVAGGIGQGHHTLGF 632
QY 170 EEVSEDMNLNLAVEYVIAELPFFIGAPDILE 199
DB 633 DVVGTMDLGDATGLARGDL-----GATDVVE 658

RESULT 15

AAW25038
ID AAW25038 standard; protein; 2329 AA.

XX AAW25038;

XX 02-FEB-1998 (first entry)

XX Partial BRCA2 cancer susceptibility gene protein.

XX BRCA2 cancer susceptibility gene; breast cancer; ovarian cancer;
XX gene therapy; prostate cancer; colorectal cancer; ocular melanoma;
XX leukaemia; human.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1267 /note= "Encoded by GA"

FT Misc-difference 1268 /note= "Encoded by AA"

XX GB2307477-A.

XX 28-MAY-1997.

XX 25-NOV-1996; 96GB-00024453.

XX 23-NOV-1995; 95GB-00023959.

XX 14-DEC-1995; 95GB-00025555.

XX 28-AUG-1996; 96GB-00017961.

XX (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:27:20 ; Search time 21.4268 Seconds
(without alignments)
776.868 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNEMTGVTTERRSVHFIAEA.....YLIVQEAQMLSEKITSEG 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	7.5	1554	4	US-09-252-991A-26814
2	95.5	7.4	2329	3	US-08-755-587-16
3	94.5	7.3	3418	2	US-08-639-501-2
4	94.5	7.3	3418	2	US-08-603-753D-4
5	94.5	7.3	3418	3	US-09-044-946-2
6	94.5	7.3	3418	3	US-08-755-587-44
7	94.5	7.3	3418	3	US-09-044-908-2
8	94.5	7.3	3418	3	US-09-089-753-4
9	94.5	7.3	3418	3	US-08-986-106-4
10	90	7.0	198	4	US-09-981-087A-25
11	90	7.0	198	4	US-09-978-382A-25
12	90	7.0	534	3	US-08-755-587-14
13	89.5	6.9	679	4	US-09-252-991A-28950
14	88	6.8	538	4	US-09-252-991A-22427
15	87	6.7	1021	4	US-09-252-991A-19205
16	84	6.5	819	3	US-09-651-656-15
17	84	6.5	819	3	US-09-650-855-15
18	84	6.5	880	4	US-09-252-991A-25046
19	82.5	6.4	394	2	US-08-646-590B-40
20	82.5	6.4	394	3	US-09-412-184-40
21	81.5	6.3	263	4	US-09-107-532A-6895
22	81.5	6.3	463	4	US-09-270-767-45923
23	81	6.3	420	4	US-09-583-110-5043
24	80.5	6.2	308	4	US-09-270-767-32667
25	80.5	6.2	308	4	US-09-270-767-47884
26	80	6.2	502	3	US-09-342-647-4
27	80	6.2	667	3	US-09-342-647-28

28	80	6.2	700	4	US-09-621-816B-2	Sequence 2, Appli
29	79.5	6.2	521	4	US-09-252-991A-25091	Sequence 25091, A
30	79.5	6.2	569	4	US-09-252-991A-29263	Sequence 29263, A
31	79.5	6.2	1206	4	US-09-245-928A-19	Sequence 19, Appl
32	79.5	6.2	1243	4	US-09-245-928A-17	Sequence 17, Appl
33	79.5	6.2	1257	3	US-08-947-823-5	Sequence 5, Appli
34	79.5	6.2	1257	4	US-09-245-928A-18	Sequence 18, Appl
35	78.5	6.1	249	4	US-09-543-681A-6793	Sequence 6793, Ap
36	78	6.1	487	1	US-08-452-722-7	Sequence 7, Appli
37	78	6.1	487	1	US-08-404-731A-7	Sequence 7, Appli
38	78	6.1	487	2	US-08-344-227-7	Sequence 7, Appli
39	78	6.1	487	2	US-08-503-226B-7	Sequence 7, Appli
40	78	6.1	487	3	US-08-721-458B-7	Sequence 7, Appli
41	77.5	6.0	289	4	US-09-395-674B-2	Sequence 2, Appli
42	77.5	6.0	363	3	US-08-984-618-13	Sequence 13, Appl
43	77.5	6.0	743	4	US-09-248-796A-15523	Sequence 15523, A
44	77	6.0	260	4	US-09-489-039A-12019	Sequence 12019, A
45	77	6.0	380	4	US-09-583-110-4774	Sequence 4774, Ap

ALIGNMENTS

RESULT 1
US-09-252-991A-26814
; Sequence 26814, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26814
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26814

Query Match 7.5%; Score 97; DB 4; Length 1554;

Best Local Similarity 19.4%; Pred. No. 0.08;
Matches 64; Conservative 33; Mismatches 95; Indels 138; Gaps 14;

Qy	1	MNEMTGVTTERRSVHFIAEALTEHCRIFFERRHV-----LVGISPFNSRFSYIVRLI	55
Db	336	LGDAATDAATHADLHFVVAQAERALSLETAHIGLEDDVQGLLFLAHVLEL-VFQLA	394
Qy	56	GWAKAQKSVSVLLAGH-EAANLLEALGTPGKAEKVRK-----EVSRRNR--FAER--	105
Db	395	GMGTROPHLTELALTEQGFAGLL--LVGDHGLVLTGIRRTVQAEGLDRRRAGFLDLA	452
Qy	106	ALVAGGDPKAIHT-----	119
Db	453	VLVEHGADTTVDADQHHVALAQGTVDQHGHHRAASLVQARLDHHAATARGRRSGFOPEH	512
Qy	120	-----FSDPID-----NKAYQLLRQVEHA-----	145
Db	513	FRLOQDRFEQFIDASTELRRNRDEGSVAAPLFRDHVSGQAVLDVVVRVGLGLVDLVRDH	572
Qy	146	FRUA-----CLDMSREAIIGR-----ARGV-----	169
Db	573	DRHARRFVLHGLGLRHDHVDHDDHIGRLGATGTHRGGRGVAGGIQEGHHATLGF	632
Qy	170	EEVSEDMNLAVEYVIAELPFFIGAPDILE	199
Db	633	DVVGTMGLGATGLARGDL-----GATDVE	658

RESULT 2
 US-08-755-587-16
 ; Sequence 16, Application US/08755587
 ; Patent No. 6045997
 ; GENERAL INFORMATION:
 ; APPLICANT: Futreal, Phillip A
 ; APPLICANT: Wooster, Richard F
 ; APPLICANT: Ashworth, Alan
 ; APPLICANT: Stratton, Michael R
 ; TITLE OF INVENTION: Materials and methods relating to the
 ; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
 ; TITLE OF INVENTION: susceptibility gene and uses thereof.
 ; NUMBER OF SEQUENCES: 222
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell Seltzer Park & Gibson
 ; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
 ; CITY: Raleigh
 ; STATE: NC
 ; COUNTRY: USA
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/755,587
 ; FILING DATE: 25-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9523959.6
 ; FILING DATE: 23-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9525555.0
 ; FILING DATE: 14-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9617961.9
 ; FILING DATE: 28-AUG-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kenneth D Sibley
 ; REGISTRATION NUMBER: 31,665
 ; REFERENCE/DOCKET NUMBER: 5405-135
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2329 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-755-587-16

Query Match 7.4%; Score 95.5; DB 3; Length 2329;
 Best Local Similarity 18.4%; Pred. No. 0.23;
 Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;
 QY 8 VTERRSVHFAEALTENCRIEERRRHV-----LVGISPFNSRSEDIYVRLIG 56
 DB 1828 VDKRNPCHVCNSEMEKTCSEFKLSNNLNVEGSSNNHISIKVSPYLSQFQDKQQLVLG 1887
 QY 57 WAKAQFQSVSVLLAGHEAANLLEALGTPRGKAERKVRKVRNRFAERALVAHGGDPKA 116
 DB 1888 TKVSLVENIHVL--GKEQA-----SPKNVMEIGKTETFS-----VPKNTNIEV 1930
 QY 117 IHTFS-----DFIDNKAYQLLRQVE-----HAFPEQPHRIACLDMSREAI 158
 DB 1931 CSTYSKDSSENYFETEAVEIAKAFMEDDELTDKLPSSHATSLFTCPENEEMLNSR--- 1987
 QY 159 IGRARGVSLAM---EEVSEDMN---LAVEVIAELPFFIGAPDILEVEETLLAYH--- 208
 DB 1988 IGRGRGEPILVUGBPSIKRNLNEDFDRIENQEKSLKASKSTPDGTIKDRRLFVHVSLE 2047
 QY 209 -----RPWKLGKISNHEFSICMRPNQGYLIIVQEMAQLSEKRIITS 249
 DB 2048 PITCVPRFTTKERQEIQNPFT---APQGBFLSKSHLYEHLTKSSS 2092

RESULT 3
 US-08-639-501-2
 ; Sequence 2, Application US/086399501
 ; Patent No. 5837492
 ; GENERAL INFORMATION:
 ; APPLICANT: Tavitgian, Sean V.
 ; APPLICANT: Kamb, Alexander
 ; APPLICANT: Simard, Jacques
 ; APPLICANT: Couch, Fergus
 ; APPLICANT: Rommens, Johanna
 ; APPLICANT: Weber, Barbara
 ; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
 ; TITLE OF INVENTION: Susceptibility Gene
 ; NUMBER OF SEQUENCES: 124
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
 ; STREET: 1201 New York Avenue N.W., Suite 1001
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 22204
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/639,501
 ; FILING DATE: 29-APR-1996
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/585,391
 ; FILING DATE: 11-JAN-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/576,559
 ; FILING DATE: 21-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/575,359
 ; FILING DATE: 20-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/573,779
 ; FILING DATE: 18-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ihnen, Jeffrey L.
 ; REGISTRATION NUMBER: 28,957
 ; REFERENCE/DOCKET NUMBER: 24884-116802-04
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-962-4810
 ; TELEFAX: 202-962-8300
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3418 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-639-501-2

Query Match 7.3%; Score 94.5; DB 2; Length 3418;
 Best Local Similarity 18.4%; Pred. No. 0.53;
 Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;
 QY 8 VTERRSVHFAEALTENCRIEERRRHV-----LVGISPFNSRSEDIYVRLIG 56
 DB 2109 VDKRNPCHVCNSEMEKTCSEFKLSNNLNVEGSSNNHISIKVSPYLSQFQDKQQLVLG 2168
 QY 57 WAKAQFQSVSVLLAGHEAANLLEALGTPRGKAERKVRKVRNRFAERALVAHGGDPKA 116
 DB 2169 TKVSLVENIHVL--GKEQA-----SPKNVMEIGKTETFS-----VPKNTNIEV 2211
 QY 117 IHTFS-----DFIDNKAYQLLRQVE-----HAFPEQPHRIACLDMSREAI 158
 DB 2212 CSTYSKDSSENYFETEAVEIAKAFMEDDELTDKLPSSHATSLFTCPENEEMLNSR--- 2268

Qy 159 IGRARGVSLMM---EEVSEDMLN---LAVEVIVIAELPFFGAPDILEVEETLLAYH--- 208
Db 2269 IGRGEPILVAGPSPKRLNLLNEFDRIENQEKSLKASKSTPDGTTKDRRLFMHVSLE 2328
Qy 209 -----RPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLSEKRIITS 249
Db 2329 PITCVPRTTKQEIQNPFT---APQBEFLSKSHLYEHLTLEKSSS 2373

RESULT 4
US-08-603-753D-4
; Sequence 4, Application US/08603753D
; Patent No. 5891857
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARLES A. TAYLOR, JR.
; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
; STREET: BOULEVARD
; CITY: DURHAM
; STATE: NORTH CAROLINA
; COUNTRY: USA
; ZIP: 27707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 KB storage
; COMPUTER: IBM PC/XT/AT compatible
; OPERATING SYSTEM: Windows 3.1
; SOFTWARE: WORD PERFECT 6.1 and ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/603,753D
; FILING DATE: 20 FEB 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/373,799
; FILING DATE: 17 JAN 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ARLES A. TAYLOR, JR.
; REGISTRATION NUMBER: 39,395
; REFERENCE/DOCKET NUMBER: 1242/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 493-8000
; TELEFAX: (919) 419-0383
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEetical: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens, sapiens
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: female breast
; CELL TYPE: normal breast tissue
; CELL LINE: HMEC
; ORGANELLE: no
; FEATURE:

; NAME/KEY: BRCA2 protein
; LOCATION: 1 to 3418; Genbank locus HSU43746
; IDENTIFICATION METHOD:
; OTHER INFORMATION: BRCA2 protein has a negative
; OTHER INFORMATION: regulatory effect on growth of human mammary cells.
; PUBLICATION INFORMATION:
; AUTHORS: Wooster, R. et al.
; TITLE: Identification of the breast cancer
; TITLE: susceptibility gene BRCA2
; JOURNAL: Nature
; VOLUME: 379
; PAGES: 789-792
; DATE: 1995
; RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
; RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
; US-08-603-753D-4

Query Match 7.3%; Score 94.5; DB 2; Length 3418;
Best Local Similarity 18.4%; Pred. No. 0.53; Mismatches 109; Indels 69; Gaps 11;
Matches 53; Conservative

Qy 8 VTERRSVHFTAEALTENCREFERRHV-----LVGISPFNSRFSEDIYRLIG 56
Db 2109 VDKRNPCHVNSEMEKTSKEFKLSNNLVGGSSNNHHSIKVSPYLSQFQDKQQLVLG 2168
Qy 57 WAKAQFASVSVLLAGHAANLLEALGTPRGKAEKVKVSRNRPAERLVAHGSDPKA 116
Db 2169 TKVSLVENIHLV--GKEQA-----SPQNVKWEIGKTTFTSD---VPVKTNI 2211
Qy 117 IHTFS---DFIDNKAYQLLRQVE-----HAFEQPHFRHACLDMSEAI 158
Db 2212 CSTYSKDSNYFETEAVEIAKAFMEDDELDTSLPSHATHSLFTCPENEEMVLNSR--- 2268
Qy 159 IGRARGVSLMM---EEVSEDMLN---LAVEVIVIAELPFFGAPDILEVEETLLAYH--- 208
Db 2269 IGRGEPILVAGPSPKRLNLLNEFDRIENQEKSLKASKSTPDGTTKDRRLFMHVSLE 2328
Qy 209 -----RPWKLGEKISNHEFSICMRPNQGYLIVQEMAQMLSEKRIITS 249
Db 2329 PITCVPRTTKQEIQNPFT---APQBEFLSKSHLYEHLTLEKSSS 2373

RESULT 5
US-09-044-946-2
; Sequence 2, Application US/09044946
; Patent No. 6033857
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/044,946
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/639,501
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/576,559
 FILING DATE: 21-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/575,359
 FILING DATE: 20-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/573,779
 FILING DATE: 18-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Ihnen, Jeffrey L.
 REGISTRATION NUMBER: 28,957
 REFERENCE/DOCKET NUMBER: 24884-116802-04
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-962-4810
 TELEFAX: 202-962-8300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-044-946-2

Query Match	7.3%;	Score	94.5;	DB	3;	Length	3418;
Best Local Similarity	18.4%;	Pred. No.	0.53;				
Matches	53;	Conservative	57;	Mismatches	109;	Indels	69;
Gaps	11;						
QY	8	VTERRSVHFTAEALTENCRIEIFFRRHV-----LVGISPFNRSFSEDIYIRLIG	56				
Db	2109	VDRKNPEHCNVSEMEKTCSEKFKLSNNLVEGGSSNNHSHSIKVSPLYLSFQDDKQQLVLG	2168				
QY	57	WAKAQFQSVSVLLAGHEAANLLEALGTPRGKAEKVKYKESRRNRRFAERLVAHGGDPKA	116				
Db	2169	TKVSLVENIHVL---GREQA-----SPKNVKEIGKTEFTFD---VPVKTNIEV	2211				
QY	117	IHTFS-----DFIDNKAYQLLRQVE-----HAFTEQPHFRHACLDMRSREAI	159				
Db	2212	CSTVSKDSSENYFETAEVIAKAFMEDDELTDLSKLPASHATSHLFTCPENEEMVLSNSR---	2268				
QY	159	IGRARGVSLMM---EEVSEDMIN---LAVEVVIAEPLFFICAPDILEVEETLLAYH----	208				
Db	2269	IGRRGEPFLILVGEPSIKRNLNNEFDRIIENQEKSLKASKSTPDGTTKDRRLFMHVSLSE	2328				
QY	209	-----RPWKLGEKISNHSEFSCIMRPNQGYLIVQEMAQMLSEKRITS	249				
Db	2329	PITCVPPRTTKEROEIQNPFT---APGOEFLSKSHLYEHLTLKSSS	2373				

```

RESULT 6
US-08-755-587-44
; Sequence 44, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
;
; PRIOR APPLICATION DATA:
;
; PRIOR APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Kenneth D Sibley
;
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
;
; INFORMATION FOR SEQ ID NO: 44:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 3418 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; US-08-755-587-44

```

Query Match	7.3%;	Score	94.5;	DB	3;	Length	3418;
Best Local Similarity	18.4%;	Pred.	No.	0.53;			
Matches	53;	Conservative	57;	Mismatches	109;	Indels	69; Gaps 11;
Qy	8	VTRRSVHFIAEALTCNCREIFERRHV-----LVGISPPNSRSESDYIVRLIG	56				
Db	2109	VDKRNPEHCNVSEMENTCKEFKLSNNLVEGSSENNHSIKVSPYLSPQQDQQLVLG	2168				
Qy	57	WAKAQPKSVSLLAGHEAANLLCALCTPRGKAERKYRKVERRRRAERALVAHGGDPKA	116				
Db	2169	TKVSIVENIHL--GKEQA-----SPKNVMELCKTETPSD---VPVKTNIEV	2211				
Qy	117	IHTFS----DFIDNKAYQLLRQVE-----HAFEPQHFRHACLDMSREAI	158				
Db	2212	CSTYSKDSENYPETEAVEATAKAPMEDDELTDSKLPGHATHSLFTCPENEEMVLSNR---	2268				
Qy	159	IGRGARGVSLMM--BVSEDMLN---LAVRYVIAELPFFIGADPILEVEETLLAYH---	208				
Db	2269	IGKRGEPLILVGEPISKINLNLEFPDRIENQESKLKASKSTDGTGDIKDRLFLMHVLSLE	2328				
Qy	209	-----RPWKLGKISNHEFSCIMPNPGYLIVQEWAOMLSKRTIS	249				
Db	2329	PITCVPRFTTKRQEQIQNPNT---APGOEFFLSKHLYEHLETLKSSS	2373				

```

RESULT 7
US-09-044-908-2
; Sequence 2, Application US/09044908
; Patent No. 6124104
; GENERAL INFORMATION:
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:

```

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/044,908
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/639,501
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/576,559
;; FILING DATE: 21-DEC-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/575,359
;; FILING DATE: 20-DEC-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/573,779
;; FILING DATE: 18-DEC-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Ihmen, Jeffrey L.
;; REGISTRATION NUMBER: 28,957
;; REFERENCE/DOCKET NUMBER: 24884-116802-04
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-962-4810
;; TELEFAX: 202-962-8300
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3418 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-044-908-2

Query Match 7.3%; Score 94.5; DB 3; Length 3418;
Best Local Similarity 10.4%; Pred. No. 0.53; 109; Indels 69; Gaps 11;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;

QY 8 VTERSVHFTAEALTECNREIFERRHV-----LVGISPFNSRFSSEDIYRLIG 56
DB 2109 VDKRNPCHCNSEMEKTCSEFKLSNNLNVEGSSNNHSIKVSPYLSQFQDQKQVLVG 2168
QY 57 WAKAQKSVSVLLAGFAANLLEALGTGRGAERKVRKVRNRFAERLVAHGGDPKA 116
DB 2169 TKVSLVNIHVL--GKQQA-----SPKNVMEIGKTTETSD---VPVKTNIIEV 2211
QY 117 IHTFS-----DPIDKAVQLLRQVE-----HAFPEQPHRACLDMSEAI 150
DB 2212 CSTYSKDSSENYFTEAVEIAKAFMEDDELTDKSLPASHATSLFTCPENEEMVLSNSR--- 2268
QY 159 IGRARGVSLAM---EEVSEMDLN---LAVEYVIAELPFFIGAPDILEVEETLLAYH--- 208
DB 2269 IGRGRGPLILVGPSPKRNLLNEFDRIENQEKSLKASKSTPDGTTKDRRLFMHVLSLE 2328
QY 209 -----RPWKLGSKNHFSICWRPNQGYLIVQEMAQMLSEKRITS 249
DB 2329 PITCVPTTKERQEIQNPNFT---APQEFELSKSLHYHLTLKSSS 2373

RESULT 8
US-09-099-753-4
; Sequence 4, Application US/09099753
; Patent No. 6149903
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.

;; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
;; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
;; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
;; NUMBER OF SEQUENCES: 29
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: ARLES A. TAYLOR, JR.
;; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
;; STREET: BOULEVARD
;; CITY: DURHAM
;; STATE: NORTH CAROLINA
;; COUNTRY: USA
;; ZIP: 27707
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
;; COMPUTER: IBM PC/XT/AT compatible
;; OPERATING SYSTEM: Windows 3.1
;; SOFTWARE: WORD PERFECT 6.1 and ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/099,753
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/603,753
;; FILING DATE: 20 FEB 1996
;; APPLICATION NUMBER: U.S. 08/373,799
;; FILING DATE: 17 JAN 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: ARLES A. TAYLOR, JR.
;; REGISTRATION NUMBER: 39,395
;; REFERENCE/DOCKET NUMBER: 1242/2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919) 493-8000
;; TELEFAX: (919) 419-0383
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3418
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: no
;; ANTI-SENSE: no
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens sapiens
;; INDIVIDUAL ISOLATE:
;; DEVELOPMENTAL STAGE: adult
;; TISSUE TYPE: female breast
;; CELL TYPE: normal breast tissue
;; CELL LINE: HMEC
;; ORGANELLE: no
;; FEATURE:
;; NAME/KEY: BRCA2 protein
;; LOCATION: 1 to 3418; Genbank locus HSU43746
;; IDENTIFICATION METHOD:
;; OTHER INFORMATION: BRCA2 protein has a negative
;; OTHER INFORMATION: regulatory effect on growth of human mammary cells.
;; PUBLICATION INFORMATION:
;; AUTHORS: Wooster, R. et al.
;; TITLE: Identification of the breast cancer
;; TITLE: susceptibility gene BRCA2
;; JOURNAL: Nature
;; VOLUME: 379
;; PAGES: 789-792
;; DATE: 1995
;; RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
;; RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-09-099-753-4

Query Match 7.3%; Score 94.5; DB 3; Length 3418;
Best Local Similarity 10.4%; Pred. No. 0.53; 109; Indels 69; Gaps 11;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;

Qy	8	VTERRSVHFTAEALTENCRIFFERRHV-----LVGSPFNRSRSESDIYIELIG	56
Db	2109	VDRNPNCHVNSEMEKTCSEKFEKLSLNVEGGSENHISIKVSPYLSQDQKQQLVLG	2168
Qy	57	WAKAQFKSVSYLLAGHAANLLEALGTPRGAERKVRKEVSRNRRFAERLVAHGGDPKA	116
Db	2169	TKVSLVENIHVL--GKEQA-----SPKVKWEICKTETFFSD---VPVKTNIEV	2211
Qy	117	IHTFS-----DFIDNKAVOLLRQVE-----HAFPEQPHFRHACLDMSREAI	158
Db	2212	CSTYSKDSSENYFETEAVEIAKAFMEDDELTDKLPSPHATHSLFTCPENBEMLNSR---	2368
Qy	159	IGRARGVSLMW--EEVSEMDLN--LAVEYVIAELPFFICAPDILEVEETLLAYH----	208
Db	2269	IGRGEPLILVGEPSIKRNLLNFEFDRIENQEKSLKASKSTDDGTTIKRRLFMHVSLSE	2328
Qy	209	-----RPWLKGEKISNHFSSICMRPNQGYLIVQEMAQMLSEKRIITS	249
Db	2329	PITCVPRTTTKEROEIQNPNT---AGQBFLSKSHLYEHLTLEKSSS	2373

RESULT 9
 US-08-986-106-4
 ; Sequence 4, Application US/08986106
 ; Patent No. 6177410
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: HOLT, JEFFREY T.
 ; APPLICANT: JENSEN, ROY A.
 ; APPLICANT: KING, MARY-CLAIRE
 ; APPLICANT: STEINER, MITCHELL S.
 ; APPLICANT: ROBINSON-BENION, CHERYL L.
 ; APPLICANT: THOMPSON, MARILYN E.
 ;
 ; TITLE OF INVENTION: THERAPEUTIC METHODS FOR
 ; TITLE OF INVENTION: PROSTATE CANCER
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ARLES A. TAYLOR, JR.
 ; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
 ; STREET: BOULEVARD
 ; CITY: DURHAM
 ; STATE: NORTH CAROLINA
 ; COUNTRY: USA
 ; ZIP: 27707
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
 ; COMPUTER: IBM PC/XT/AT compatible
 ; OPERATING SYSTEM: Windows 3.1
 ; SOFTWARE: WORD PERFECT 6.1 and ASCII
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/986,106
 ; FILING DATE:
 ; CLASSIFICATION:
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/603,753
 ; FILING DATE: 20 FEB 1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ARLES A. TAYLOR, JR.
 ;
 ; REGISTRATION NUMBER: 39,395
 ;
 ; REFERENCE/DOCKET NUMBER: 1242/3
 ;
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (919) 493-8000
 ; TELEFAX: (919) 419-0383
 ;
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3418
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ;
 ; FEATURE:
 ; NAME/KEY: BRCA2 protein
 ; PUBLICATION INFORMATION:
 ; AUTHORS: Wooster, R. et al.
 ; TITLE: Identification of the breast cancer

```

; TITLE: susceptibility gene BRCA2
; JOURNAL: Nature
; VOLUME: 379
; PAGES: 789-792
; DATE: 1995
; RELEVANT RESIDUES IN SEQ ID NO: 4: granin box domain at
; RELEVANT RESIDUES IN SEQ ID NO: amino acids 3334-3344
; US-08-986-106-4

      7.3%; Score 94.5; DB 3; Length 3418;
      Best Local Similarity 18.4%; Pred. No. 0.53;
      Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;

Qy 8 VTRRSVHFAETALTENCREIFEIRRHV-----LVGISPPNSRFSEDIYRLIG 56
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2109 VDKRNPEHCNVSEWXTCSKEFKLSNLLNVEGSSENNHSIKVSPYLSQFQDKQLVLG 2168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 57 WAKAQPKSVSVLLAGHAANLLALGTTPGKAERKVRKESNRNRAERALVAHGGDPKA 116
   | : : | | | | | | | | | | | | | | | | | | | | | | | |
Db 2169 TKVSLVENIHVL--GREQA-----SPKNVKEIGKTETPSD---VPVKTNIEV 2211
   | : : | | | | | | | | | | | | | | | | | | | | | | | |
Qy 117 IHTFS---DFIDNKAYQLLRQVE-----HAPFQQPHFRHACLDMSREAL 158
   | : : | | | | | | | | | | | | | | | | | | | | | | | |
Db 2212 CSTSYKDSSENYFETAEVAKAFMEDDELTDKLPASHATHSLTCTENEEMVLSNR 2268
   | : : | | | | | | | | | | | | | | | | | | | | | | | |
Qy 159 IGRARGVSLMM---EVSSEDLN---LAVEYVIAELPFFTGPADILEVEETLLAYH--- 208
   | : : | | | | | | | | | | | | | | | | | | | | | | | |
Db 2269 IGRERGEPLLVGEPISIKRNLLNEFDRIIENQESLKASKSTPDGIIKORRFLFMHVSL 2328
   | : : | | | | | | | | | | | | | | | | | | | | | | | |
Qy 209 -----RPWKLGKTSINHEFSCMRPNQGYLIVQEMAQMLSEKRITS 249
   | : : | | | | | | | | | | | | | | | | | | | | | | | |
Db 2329 PITCVPRFTTKRQETQNPNFT---APGQEFLSKSHLYEHLITLSEKSS 2373
   | : : | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 10
US-09-981-087A-25
Sequence 25, Application US/09981087A
Patent No. 6768042
GENERAL INFORMATION:
APPLICANT: Yanofsky, Martin F.
APPLICANT: Liljegren, Sarah
APPLICANT: Farrandiz, Cristina
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
TITLE OF INVENTION: Plants
FILE REFERENCE: 19452A-000940US
CURRENT APPLICATION NUMBER: US/09/981,087A
CURRENT FILING DATE: 2002-05-21
PRIOR APPLICATION NUMBER: US 60/090,649
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: US 09/339,998
PRIOR FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 198
TYPE: PRT
ORGANISM: Arabidopsis sp.
US-09-981-087A-25

	Query Match	7.0%;	Score 90;	DB 4;	Length 198;
	Best Local Similarity	29.3%;	Pred. No.	0.023;	
	Matches	39;	Conservative	22;	Mismatches 46; Indels 26; Gaps 8;
Qy	97	SRNRRFAERALVANGDPPKAIHTFSDFIDNKAYQLLRQVEHAFFPEOPHFRACLDMSRE	156		
		: :: :			
Dd	20	SRSNHSPKPSMM---EPQPHLLMDW--NKANDLLTQE-HAFLNDPH--HLMLDPPE	70		
		: :: :			
Qy	157	AATGRARGYSLMEEVSDEMLNL-AYEYVIAEL-----PFFICAPD-----LLEVEETL	204		
		: :: :			
Dd	71	TLIHLED-----EYEDDMAMKEQMVIANQPVDIDPATVPKNRRNRVRISDDQTV	125		
Ov	205	LAYHRPWKLGEKI	217		


```
Db 449 REAFHLVVEQLDNRLEIGRGEDVDHIAHTESRTGEIHV-----AGVLQA 496
QY 71 GHEAAN--LLEALGTRGKAERKVRKEVSNRRFAERLVAHGGDPKAIHTFSDFDNKA 128
Db 497 GQSPQOQFALVDV-----AAVQVDHLQVGRGITEAVDGRHGGDDDRINTLEQRIGRRQ 550
QY 129 YQLRQVEHAFFEQPHFRHACLDMSREALIGRAR-GVSLMMEEVSEDMNLNLAVEYVIAE 187
Db 551 AHLLDMVD-----RGVLIDIG-----IGRDVGLRLVVVVGVNEVLHGIVREERLE 597
QY 188 LPFFIGAPDILEVETLLAYHRPWKLGKYSNHE-PSICMRPNQG 231
Db 598 LPVELGRQGLVRRQDQGRALH-----LGNHVGDASLARARHPQOG 638

RESULT 14
US-09-252-991A-22427
; Sequence 22427, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22427
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22427
```

```
Query Match 6.8%; Score 88; DB 4; Length 538;
Best Local Similarity 26.0%; Pred. No. 0.19;
Matches 44; Conservative 22; Mismatches 55; Indels 48; Gaps 8;

QY 60 AQFKSVSVLLAGHEAAN--LLE-----ALGTPRG-----KAERKVRKEVSNRRF----- 102
Db 92 AQFAYVARPVVGHQAGGGTVFERGAVEAQRGLLEEVFEQQQDVLAAVAQRQFQAHYF 151
QY 103 -----AERLVAHGG-----DPKAIH-----TFSDFDNKAQQLLRQVEHA- 139
Db 152 EPVVKVAAELAAALAHGGEVCLGRRDHPAVHRDLRVRAEAFQGVFLQHPQLDLQVDRHAL 211
QY 140 -FFEQPHFRHACLDMSREALIGRARGVSLMMEEVSEDMNLNLAVEYVIAE 187
Db 212 DLVEEQRAAVGMLDLPPLVGTGEGVRLVAE-----YLAVEQVLRQ 253
```

```
RESULT 15
US-09-252-991A-19205
; Sequence 19205, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19205
; LENGTH: 1021
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19205

Query Match 6.7%; Score 87; DB 4; Length 1021;
Best Local Similarity 25.7%; Pred. No. 0.65;
Matches 45; Conservative 31; Mismatches 67; Indels 32; Gaps 8;

QY 26 REIFERRRHV-----LVGISPFNSRFSYIVRLIGWAKAQFKSVSVLLAGHEAANLL-E 79
Db 586 RDVEERLRHLHYDELGTGIA--NRRLFRERLHEAQAQRQEGGLALLLLDLRFKLLND 643
QY 80 ALGTPRGKAERKVRKEVSNRRRFAERA--LVAHGGDPKAI---HTFSDFDNKAQQLLR 133
Db 644 SLG--HEVADQLLRMAQRLRRTAPEANTLARLAGDEFAILLDGGGTGAALSRLAERLLV 701
QY 134 QVEHAFPEQPHFRHACLDMSREALIGRARGVSLMMEEVSE-----DMLNLAVEY 183
Db 702 Q-----LRQPVSVLGHHELILGASIGISLFSQAREISVLMQANLMAQH 745
```

Search completed: December 10, 2004, 13:46:53
Job time : 23.4268 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:45:11 ; Search time 73.4634 Seconds
(without alignments)
1220.359 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNEMTGMVTRRSHVHFAIA.....YLIVQVMAQMLSEKRTISEG 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1585576 seqs, 357178320 residues
Total number of hits satisfying chosen parameters: 1585576

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first: 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pdb*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pdb*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pdb*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pdb*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pdb*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pdb*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pdb*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pdb*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pdb*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pdb*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pdb*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pdb*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pdb*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pdb*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pdb*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pdb*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pdb*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pdb*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1289	100.0	251	15	US-10-627-124-8
2	102.5	8.0	557	9	Sequence 8, Appli
3	102.5	8.0	557	11	Sequence 60, Appl
4	102.5	8.0	557	14	Sequence 60, Appl
5	100	7.8	1093	14	Sequence 14825, A
6	96.5	7.5	1710	16	Sequence 149352, A
7	94.5	7.3	3418	14	Sequence 42, Appl
8	94.5	7.3	3418	15	Sequence 1, Appli
9	93.5	7.3	495	16	Sequence 114455, A
10	93.5	7.3	3418	16	Sequence 178, App
11	93	7.2	1156	14	Sequence 21367, A
12	91.5	7.1	610	16	Sequence 137309, A
13	90.5	7.0	545	16	Sequence 148730, A

14	90	7.0	138	9	US-09-978-730-25	Sequence 25, Appl
15	90	7.0	138	9	US-09-978-729A-25	Sequence 25, Appl
16	90	7.0	138	9	US-09-981-087A-25	Sequence 25, Appl
17	90	7.0	138	9	US-09-978-382A-25	Sequence 25, Appl
18	90	7.0	138	10	US-09-978-740A-25	Sequence 25, Appl
19	90	7.0	200	10	US-09-934-455-128	Sequence 128, App
20	90	7.0	200	14	US-10-286-264-90	Sequence 90, Appl
21	90	7.0	200	14	US-10-225-066A-114	Sequence 114, App
22	90	7.0	200	15	US-10-374-780A-158	Sequence 158, App
23	90	7.0	200	15	US-10-412-699B-526	Sequence 526, App
24	90	7.0	1172	17	US-10-739-930-6385	Sequence 6385, Ap
25	88.5	6.9	284	15	US-10-282-122A-52738	Sequence 52738, A
26	87.5	6.8	724	14	US-10-369-493-13830	Sequence 13830, A
27	87	6.7	365	14	US-10-369-493-23225	Sequence 23225, A
28	87	6.7	977	16	US-10-437-963-124951	Sequence 124951, A
29	86.5	6.7	1024	16	US-10-437-963-124952	Sequence 124952, A
30	86.5	6.7	249	16	US-10-437-963-184335	Sequence 184335, A
31	85	6.7	179	14	US-10-365-742-74	Sequence 74, Appl
32	85.5	6.6	560	15	US-10-425-114-57637	Sequence 57637, A
33	85.5	6.6	776	17	US-10-425-115-228544	Sequence 228544, A
34	85.5	6.6	786	15	US-10-425-114-59864	Sequence 59864, A
35	85	6.6	716	14	US-10-369-493-774	Sequence 774, App
36	85	6.6	1828	16	US-10-437-963-123225	Sequence 123225, A
37	84.5	6.6	370	17	US-10-425-115-350281	Sequence 350281, A
38	84.5	6.6	679	16	US-10-437-963-185821	Sequence 185821, A
39	84.5	6.6	2033	16	US-10-408-765A-419	Sequence 419, App
40	84	6.5	419	14	US-10-369-493-16720	Sequence 16720, A
41	84	6.5	425	14	US-10-369-493-4922	Sequence 4922, Ap
42	84	6.5	425	14	US-10-369-493-7680	Sequence 7680, Ap
43	84	6.5	1013	16	US-10-437-963-199969	Sequence 199969, A
44	83.5	6.5	965	16	US-10-437-963-182421	Sequence 182421, A
45	83.5	6.5	1170	14	US-10-369-493-3006	Sequence 3006, Ap

ALIGNMENTS

RESULT 1
US-10-627-124-8
; Sequence 8, Application US/10627124
; Publication No. US20040096944A1
; GENERAL INFORMATION:
; APPLICANT: NOVOZYMES BIOTECH, INC.
; TITLE OF INVENTION: Methods For Producing Biological Substances In Pigment-Deficient
; FILE OF INVENTION: Mutants Of Bacillus Cells
; FILE REFERENCE: 10302.200-WO
; CURRENT APPLICATION NUMBER: US/10/627,124
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: US 60/398,853
; PRIOR FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 8
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-627-124-8

Query Match	100.0%	Score 1289;	DB 15;	Length 251;
Best Local Similarity	100.0%;	Pred. NO. 1e-128;		
Matches 251;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MNEMTGMVTRRSHVHFAIALEALGTGPRGAERKVRKESRNRFAERLVAHGGDPKAIHTF	60	
Db	1	MNEMTGMVTRRSHVHFAIALEALGTGPRGAERKVRKESRNRFAERLVAHGGDPKAIHTF	60	
Qy	61	QFKSVSVLLAGHAANLLEALGTGPRGAERKVRKESRNRFAERLVAHGGDPKAIHTF	120	
Db	61	QFKSVSVLLAGHAANLLEALGTGPRGAERKVRKESRNRFAERLVAHGGDPKAIHTF	120	
Qy	121	SDFIDNKAYQLLRQVEHFAFFQPHFHAACLDMSREAIIGRARGVSLMMEVSDMLNLA	180	
Db	121	SDFIDNKAYQLLRQVEHFAFFQPHFHAACLDMSREAIIGRARGVSLMMEVSDMLNLA	180	

Applicant

```
;
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-856-60

Query Match      8.0%; Score 102.5; DB 11; Length 557;
Best Local Similarity 26.3%; Pred. No. 0.1;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

Qy 84 PRGAERKVRKEVSRNRRFAERLVAHGGDPKAIHTFSDIDNKAYOLLRQEVHEHAFPEQ 143
Db 95 PEGAETRKVIETKLAR-----FVAEGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 144
Qy 144 PHERHACLDMSREAIIGRARGVSLMMEEVSEDMNLNLAVEVIAELPFI--GAPDILEVE 201
Db 145 -YYRKKVAEIRKEAQSKASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 194
Qy 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 195 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEEFKAKASSTG 237

RESULT 4
US-10-102-627-60
; Sequence 60, Application US/10102627
; Publication No. US20030054377A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ24C1
; CURRENT APPLICATION NUMBER: US/10/102,627
; CURRENT FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 110
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (86)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-627-60

Query Match      8.0%; Score 102.5; DB 14; Length 557;
Best Local Similarity 26.3%; Pred. No. 0.1;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

Qy 84 PRGAERKVRKEVSRNRRFAERLVAHGGDPKAIHTFSDIDNKAYOLLRQEVHEHAFPEQ 143
Db 95 PEGAETRKVIETKLAR-----FVAEGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 144
Qy 144 PHERHACLDMSREAIIGRARGVSLMMEEVSEDMNLNLAVEVIAELPFI--GAPDILEVE 201
Db 145 -YYRKKVAEIRKEAQSKASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 194
Qy 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 195 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEEFKAKASSTG 237

RESULT 5
US-10-156-761-14825
; Sequence 14825, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
```

```
;
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-856-60

Query Match      8.0%; Score 102.5; DB 9; Length 557;
Best Local Similarity 26.3%; Pred. No. 0.1;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

Qy 84 PRGAERKVRKEVSRNRRFAERLVAHGGDPKAIHTFSDIDNKAYOLLRQEVHEHAFPEQ 143
Db 95 PEGAETRKVIETKLAR-----FVAEGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 144
Qy 144 PHERHACLDMSREAIIGRARGVSLMMEEVSEDMNLNLAVEVIAELPFI--GAPDILEVE 201
Db 145 -YYRKKVAEIRKEAQSKASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 194
Qy 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 195 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEEFKAKASSTG 237

RESULT 3
US-09-764-856-60
; Sequence 60, Application US/09764856
; Publication No. US20040101927A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ24
; CURRENT APPLICATION NUMBER: US/09/764,856
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (86)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-856-60

Query Match      8.0%; Score 102.5; DB 9; Length 557;
Best Local Similarity 26.3%; Pred. No. 0.1;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

Qy 84 PRGAERKVRKEVSRNRRFAERLVAHGGDPKAIHTFSDIDNKAYOLLRQEVHEHAFPEQ 143
Db 95 PEGAETRKVIETKLAR-----FVAEGGPELEKVMEDYKDNPAFAFLHDKNSREFL-- 144
Qy 144 PHERHACLDMSREAIIGRARGVSLMMEEVSEDMNLNLAVEVIAELPFI--GAPDILEVE 201
Db 145 -YYRKKVAEIRKEAQSKASQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 194
Qy 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRITSEG 251
Db 195 TIALQNNR-----ENQAFSFLYEPNSQGYKYRQKLEEFKAKASSTG 237

RESULT 3
US-09-764-856-60
; Sequence 60, Application US/09764856
; Publication No. US20040101927A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ24
; CURRENT APPLICATION NUMBER: US/09/764,856
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (86)
```

```
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14825
; LENGTH: 1093
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-14825

Query Match 7.8%; Score 100; DB 14; Length 1093;
Best Local Similarity 26.5%; Pred. No. 0.49;
Matches 36; Conservative 15; Mismatches 47; Indels 38; Gaps 4;

Qy 19 EALTENC-----REIFERRHVLGISPNSRFSDDYIRLIGWAKAQ 61
Db 939 EALRHCDRAGVPYTTADGRGVVLEMYERLVEERTGLPTFKDFTD----- 986
Qy 62 FKSVSLLAGHEA-ANLE-----ALGTPGKAERKVRKVRNRRAERAAVAHGGDPK 115
Db 987 ---VSLTROHRAEDPRAERWDLVAFGTGLTAYSELTDPEQRRLTAQSLLAAGDPE 1043
Qy 116 AIHTFSDFDINKAYQL 131
Db 1044 AMELDEDFDLAEYAM 1059

RESULT 6
US-10-437-963-199352
; Sequence 199352, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazov, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 199352
; LENGTH: 1710
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_94925C.1.pap
; US-10-437-963-199352

Query Match 7.5%; Score 96.5; DB 16; Length 1710;
Best Local Similarity 24.0%; Pred. No. 2.2;
Matches 41; Conservative 26; Mismatches 41; Indels 63; Gaps 8;

Qy 54 LIGWAKQFKSVSVLLAGHEAANLEALGTPGKAERKVRKVRNRRAERAAVAHGGD 113
Db 889 LQGSVKVFNVS-----RAIEKARDLA-----TLVSHNAD 920
Qy 114 PKAHTFSDFDINKAYQ-----LLR-----QVEHAFPEQPHFRHACLDM 153

; APPLICANT: YSEIRSSDVLNELLRYEEMLWLRQSRISWLKVEKSTHEMEKIETDYFKOMFTADPCIDQ 980
; APPLICANT: SREAIIGRAGVSIIMEEVSDM-LNLAVEVY---IAELPFFIG---APDI 197
; APPLICANT: SRVS-----RLFQKVSPEMNVLDCKDFTQEOFIADALFOIGPIKAPGL 1023
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14825
; LENGTH: 1093
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-14825

Query Match 7.3%; Score 94.5; DB 14; Length 3418;
Best Local Similarity 18.4%; Pred. No. 9.6;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;

Qy 8 VTERRSVHFTAEALTENCRIEIPERRHV-----LVGISPFNSRFSDDYIRLIG 56
Db 2109 VDKRNPEHCNVSEMEKTCSEFKLSNNLNVGSSNNHISIKVSPYLSQFQODRQQLVLG 2168
Qy 57 WAKAQFKSVSVLLAGHEAANLEALGTPGKAERKVRKVRNRRAERAAVAHGGDPKA 116
Db 2169 TKVSLVENIHVL--GKEQA-----SPKNVMEIGKTETTFSD---VPVKTNIEV 2211
Qy 117 IHTFS-----DFIDNKAYQLLRQVE-----HAFPEQPHFRHACLDMREAI 158
Db 2212 CSTYKSDSENYFTEAVEIAKAFWEDDELTDKLPASHATSLFTCPNEEWLNSNR--- 2268
Qy 159 IGRAGVSLMM-----EEVSEDMAN---LAVEYIAELPFFIGADPILVEETLTAH---- 208
Db 2269 IGKRRGEPLILVGEPSIKRNLNNEFDRIIEHQESLAKSKSTPDGTIKDRRLFMHVSLE 2328
Qy 209 -----RPMKLGKISNHEFSICMRPNQCYLIVQEAQMLSEKRITS 249
Db 2329 PITCVPRFTTKERQEIQNPNFT---APQBFSLKSHLYEHLTLEKSSS 2373

RESULT 8
US-10-634-574-1
; Sequence 1, Application US/10634574
; Publication No. US20040072268A1
; GENERAL INFORMATION:
; APPLICANT: The Wistar Institute
; APPLICANT: Ramin Shiekhatar
; TITLE OF INVENTION: METHODS FOR REGULATING BRCA1-BRCA2-CONTAINING COMPLEX ACTIVITY
; FILE REFERENCE: WSTR-0014B
; CURRENT APPLICATION NUMBER: US/10/634,574
; CURRENT FILING DATE: 2003-08-05
```

	Prior Application Number:	US 60/401,433	
	Prior Filing Date:	2002-08-05	
	Prior Application Number:	US 60/449,950	
	Prior Filing Date:	2003-02-24	
	Number of Seq ID NOS:	26	
	Software:	PatentIn version 3.1	
	Seq ID NO 1		
	Length:	3418	
	Type:	PRT	
	Organism:	Homo sapiens	
	US-10-634-574-1		
	Query Match	7.3%; Score 94.5; DB 15; Length 3418;	
	Best Local Similarity	18.4%; Pred.No.9.6;	
	Matches	53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;	
QY	8 VTERS	VHFAIALTENCHEIFERRHV-----LVGISPFNSRSEDVIYLIG 56	
DB	2109 VKRN	PEHCVMSEKTSKEFKLSNNLNVEGSSENNHSIKVSPYLSQQFOODKKQLVLG 2168	
QY	57 WAKA	QPKSVSVLLAGHEAANLLEALGTGRKAERKVRSNRRAERFARALVAHGDPKA 116	
DB	2169 TKVS	LVENIHVL--GKEQA-----SPKNVKMEIKGTETFS---VPVTNIEV 2211	
QY	117 IHTF	S----DFIDNKAYOLLROVE-----HAFFEQPHFRACLDMSREAI 158	
DB	2212 CSTY	KDSENYPETEAVEITAKAFMEDDELTDKLPSSHATHSLFTCPENEEVLSNR---2268	
QY	159 IGRAR	GVSMM--BEVSDMLN---LAVYYIAELPFGAPDILEVEETLLAYH---208	
DB	2269 IGKR	RGEPLILVGPSIKRNLNEFDRIENOESKLKASKSTPDGTIKDRFLFMHHVSL 2328	
QY	209 -----RPW	LGEKISNHFSICMRPNQGYLIVQMAMLSRKITS 249	
DB	2329 PITC	VFTRTKERQEIQNPNT---APGQEFLSKHLYEHULTLEKSSS 2373	
	RESULT 9		
	US-10-437-963-114455		
	; Sequence 114455, Application US/10437963		
	; Publication No. US20040123343A1		
	; GENERAL INFORMATION:		
	; APPLICANT: La Rosa, Thomas J.		
	; APPLICANT: Kovalic, David K.		
	; APPLICANT: Zhou, Yihua		
	; APPLICANT: Cao, Yongwei		
	; APPLICANT: Wu, Wei		
	; APPLICANT: Boukharov, Andrey A.		
	; APPLICANT: Barbazuk, Brad		
	; APPLICANT: Li, Ping		
	; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With		
	; FILE REFERENCE: 38-21(53221)B		
	; CURRENT APPLICATION NUMBER: US/10/437,963		
	; CURRENT FILING DATE: 2003-05-14		
	; NUMBER OF SEQ ID NOS: 204966		
	; SEQ ID NO 114455		
	; LENGTH: 495		
	; TYPE: PRT		
	; ORGANISM: Oryza sativa		
	; FEATURE:		
	; OTHER INFORMATION: Clone ID: PAT_MRT4530_18144C.1.pep		
	US-10-437-963-114455		
	Query Match	7.3%; Score 93.5; DB 16; Length 495;	
	Best Local Similarity	22.7%; Pred.No.0.79;	
	Matches	57; Conservative 32; Mismatches 97; Indels 65; Gaps 12;	
QY	49 DTYIR	LIGHWAKAQKSVSVLLAGHEAANLLEALGTGRG-KAEKRVKE--VSRNRRFAER 105	
DB	115 DYIN	ONCHWSLLPLLALKIIFAG--LANLIYGPTVFGSRTDRLANDSGVMSLRLAAS 172	
QY	106 ALVA	HGDGPDKAITFSDFDINDKAYOLLROVEHAFEOPHFHRACL-----DNMREAII 159	


```

Db      71  TLIHLEDED-----EYDEDEDAMKENQYMTAVMQPVDIDPATVFKPNRRNVRISDDPQT 125

Qy      205  LAYHRPWKLGEKI 217
          : | : ||
Db      126  VARRRRERISEKI 138

Search completed: December 10, 2004, 14:02:31
Job time : 75.4634 secs

```

```

; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljefgren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000920US
; CURRENT APPLICATION NUMBER: US/09/978,730
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 25
; LENGTH: 198
; TYPE: PR
; ORGANISM: Arabidopsis sp.
; US-09-978-730-25

```

Query Match	7.0%;	Score 90;	DB 9;	Length 198;
Best Local Similarity	29.3%;	Pred. No. 0.51;		
Matches 39;	Conservative 22;	Mismatches 46;	Indels 26;	Gaps 8;
QY	97	SRNRRFAERALVAHGGGPKAIHTSDIDNKAYOLLQVEHAFEEOPHFRHACLDMSRE	156	
Db	20	SRNSHSPRSMM---EPQPHLLMDW--NKNDLLTQE--HAAFLNDPH--HMLDPPPE	70	
QY	157	AITGRARGVSLMBEVESEDLNL--AVEYVIAEL-----PFFIGAPD-----ILEVEETL	204	
Db	71	TLIHLDSD-----EYEDMDAMKEMQYMIAMVQPDIDPATVFKPNRRNVRISDDPQT	125	
QY	205	LAYHRPWLKGEKI	217	
Db	126	VARRRRRIRISEKI	138	

```

RESULT 15
US-09-978-729A-25
; Sequence 25, Application US/09978729A
; Patent No. US20020178465A1
; GENERAL INFORMATION:
; APPLICANT: Yanofsky, Martin F.
; APPLICANT: Liljgren, Sarah
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Selective Control of Lignin Biosynthesis in Transgenic
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 19452A-000950US
; CURRENT APPLICATION NUMBER: US/09/978,729A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/090,649
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: US 09/339,998
; PRIOR FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-978-729A-25

```

	Query Match	7.0%;	Score 90;	DB 9;	Length 198;
	Best Local Similarity	29.3%;	Pred. No. 0.51;		
	Matches	39;	Conservative 22;	Mismatches 46;	Indels 26; Gaps 8;
Qy	97	SRNRRFAERALVAHGDPKALHTSFDFIDNKAYOLLRQVEVEHAFEQPHFACILDMSR	156		
Db	20	SRSNHSPKRSNM-----EPQPHLLMDM--NKANDLLTQE-HAAFUNDPH--HLMDDPPPE	70		
Qy	157	AIIGRARGVSILMMEVSDMLNL-AVEYVIAEL-----PFFIGAPD-----ILVEVEETL	204		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2004, 11:49:51 ; Search time 3348.71 Seconds
(without alignments)
10633.698 Million cell updates/sec

Title: US-10-627-124-7

Perfect score: 753
Sequence: 1 gtaatgagatgacccgaat.....aacggatcacatctgaagga 753

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 segs, 2364489745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 10%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	753	100.0	17739	1 AF017113	AF017113 Bacillus
C 2	753	100.0	194692	1 BSUB0018	Z99121 Bacillus su
C 3	61.6	8.2	349483	1 BX571859	BX571859 Phototrab
C 4	61.6	8.2	349980	6 AX770907	AX770907 Sequence
C 5	48.6	6.5	7218	6 I66494	I66494 Sequence 14
C 6	43	5.7	175919	9 AC040918	AC040918 Homo sapi
C 7	42.8	5.7	1038	5 AF109942	AF109942 Amblyramp
C 8	42.8	5.7	1041	5 AF099355	AF099355 Amblyerc
C 9	41.8	5.6	1659	6 E08319	E08319 cDNA encodi
C 10	41.2	5.5	1006	5 AF526473	AF526473 Motacilla
C 11	41.2	5.5	1006	5 AF526481	AF526481 Motacilla
C 12	41.2	5.5	1659	6 I40164	I40164 Sequence 1
C 13	41.2	5.5	1659	6 I40165	I40165 Sequence 2
C 14	40.8	5.4	1056	5 AY294513	AY294513 Aerodramu
C 15	40.8	5.4	1064	5 AY294501	AY294501 Aerodramu
C 16	40.8	5.4	1068	5 AY294514	AY294514 Aerodramu
C 17	40.8	5.4	1072	5 AY294497	AY294497 Aerodramu
C 18	40.8	5.4	1076	5 AY294502	AY294502 Aerodramu
C 19	40.8	5.4	1076	5 AY294503	AY294503 Aerodramu

C 20	40.8	5.4	1076	5 AY294506	AY294506 Aerodramu
C 21	40.8	5.4	1076	5 AY294515	AY294515 Aerodramu
C 22	40.4	5.4	2000	6 AX655393	AX655393 Sequence
C 23	40.2	5.3	676	1 L81151	L81151 Buchnera ap
C 24	40.2	5.3	125020	9 AF429315	AF429315 Homo sapi
C 25	40	5.3	110000	1 AE017261_03	Continuation (4 of
C 26	39.8	5.3	1076	5 AY294498	AY294498 Aerodramu
C 27	39.8	5.3	2000	6 AX655393	AX655393 Sequence
C 28	39.8	5.3	221444	2 AC113313	AC113313 Mus muscu
C 29	39.8	5.3	247757	2 AC121282	AC121282 Mus muscu
C 30	39.6	5.3	1041	5 AF099341	AF099341 Icterus m
C 31	39.6	5.3	1041	5 AY049515	AY049515 Catharus
C 32	39.6	5.3	181029	10 AC140183	AC140183 Mus muscu
C 33	39.4	5.2	1038	5 AF407463	AF407463 Brachypte
C 34	39.4	5.2	1088	5 AF407030	AF407030 Chloebia
C 35	39.4	5.2	116745	2 AC146854	AC146854 Medicago
C 36	39.4	5.2	248537	2 AC124344	AC124344 Mus muscu
C 37	39.2	5.2	1076	5 AY294505	AY294505 Aerodramu
C 38	39	5.2	2776	5 BC073566	BC073566 Xenopus l
C 39	39	5.2	2846	10 BC026771	BC026771 Mus muscu
C 40	38.8	5.2	1038	5 AF109962	AF109962 Dives war
C 41	38.8	5.2	1041	5 AF099343	AF099343 Icterus m
C 42	38.6	5.1	1041	5 AF099346	AF099346 Icterus o
C 43	38.4	5.1	110000	8 CR380955_03	Continuation (4 of
C 44	38	5.0	676	1 AF058434	AF058434 Buchnera
C 45	38	5.0	1006	5 AF526483	AF526483 Motacilla

ALIGNMENTS

RESULT 1	AF017113	47739 bp	DNA	linear	BCT 23-MAY-2000
LOCUS	Bacillus subtilis	300-304	degree	genomic	sequence.
DEFINITION	AF017113				
ACCESSION	AF017113				
VERSION	AF017113.1	GI:2618830			
KEYWORDS	Bacillus subtilis				
SOURCE	Bacillus subtilis				
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
REFERENCE	1 (bases 1 to 47739)				
AUTHORS	Reizler, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stulke, J., Karamata, D., Sailer, M.H. Jr. and Hillen, W.				
TITLE	A novel protein kinase that controls carbon catabolite repression in bacteria				
JOURNAL	Mol. Microbiol.	27 (6),	1157-1169	(1998)	
MEDLINE	98230327				
PUBMED	9570401				
REFERENCE	2 (bases 30303 to 37035)				
AUTHORS	Robinson, C., Rivolta, C., Karamata, D. and Moir, A.				
TITLE	The product of the yvoC (gepF) gene of Bacillus subtilis is required for spore germination				
JOURNAL	Microbiology	144 (Pt 11),	3105-3109	(1998)	
MEDLINE	99061203				
PUBMED	9846746				
REFERENCE	3 (bases 1 to 47739)				
AUTHORS	Lazarevic, V., Soldo, B., Rivolta, C., Reynolds, S., Maue, C. and Karamata, D.				
TITLE	Nucleotide sequence of the 300-304 chromosomal segment of Bacillus subtilis				
JOURNAL	Unpublished				
PUBMED	4 bases 1 to 47739				
AUTHORS	Lazarevic, V.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-AUG-1997)				
FEATURES	Switzerland				
source	Location/Qualifiers				
	1. .47739				
	/organism="Bacillus subtilis"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1423"				
	/map="300-304 degrees"				

```
gene
/11. .1184
/gene="prfB"
RBS
/71. .77
/gene="prfB"
CDS
/join(83. .154,156. .1184)
/gene="prfB"
/notes="at low concentrations of RF-2 nucleotides 155. .158
(UGAC) are translated as Asp, Phe, H.J. Nucleic Acids Res.
20. 4423-4428, 1992"
/codon_start=1
/transl_table=11
/product="putative peptide chain release factor RF-2"
/protein_id="AAC67303.1"
/db_xref="GI:2618874"
/translation="MELSEIRAELENWASRLADFRGSLDLSEKAEIRIELDEQMDPE
FNDQKRAQVINEANGLDYVNSYKLNESHEELQMTDHLKKEPDITDLQLEKEKL
KSTLRFNFEPLQLLSYFDKNNAILLEHPGAGGTESQDQWGSMLLRYTRWGERGF
KVTLLDLPQDEAGIKSVTLIKGHNAVGYLKAEGVHRLVRIISPDSSGRRTSFSV
CEVMPFNDEIDIDRTEDIKVDTRYASGAGGQHVNTDSAVRIIHLPTNVVTCOTE
RSQIKRERAMKAKLYQRRIEEOQAEDELIRGEQKEIGWSQIRSYVFPYSMVK
DHRNTMGVQAVMDGDIDTDFIDAYLRSLKLS"
143. .149
/gene="prfB"
RBS
1338. .2195
/gene="yvjA"
CDS
1338. .1342
/gene="yvjA"
1350. .2195
/gene="yvjA"
/notes="similar to thermophilic bacterium PS-3 hypothetical
protein: PIR Accession Number S43727 and to Bacillus
subtilis Yqfu, YpJc and YitB proteins"
/codon_start=1
/transl_table=11
/product="yvjA"
/protein_id="AAC67260.1"
/db_xref="GI:2618831"
/translation="MDVRNKTILRDYVYILIGAAITAVSNVFLPNKIAAGVSG
ISTILQSYGEAAYVOMIINIPFIAGVILLGKFGKLTAGSVFLPVFLFRDIQF
ATHLELLAIFGVGIGIGIVYLGKSGTGSTALAAQIHKYSGLSLKGKCLAIIDGM
IVVTAMVFNIEQLYAMLGVVYSKTDVVOVGFNRSKMALIITKQEQAKBAVLQK
IDRGVTKISAGVGTDDDRPILMVCVVGQTEFTKLQIKVQIDESAFAVIVADASEVLGE
GFKRA"
2223. .2582
/gene="cccB"
RBS
2223. .2229
/gene="cccB"
CDS
2244. .2582
/gene="cccB"
/notes="putative lipoprotein"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="cytochrome c-551"
/protein_id="AAC67261.1"
/db_xref="GI:2618832"
/translation="MKSILMIGFALSVLAAACGSNDKBEKTDGSKTEATASEG
EELYQSCVCGCHQKLEGVSGPNLQEVGGYDHEKHIESIIKNGRGNMPLGLVDNEAA
VIAKWASEKK"
2806. .3502
/gene="ftsE"
RBS
2806. .2813
/gene="ftsE"
CDS
2816. .3502
/gene="ftsE"
/codon_start=1
/transl_table=11
/product="cell division ATP-binding protein"
/protein_id="AAC67262.1"
/db_xref="GI:2618833"
/translation="MIEMKEVRYKAYPNGVKALNGISVTHHPGFVYVVGSGAGKSTFF
IKMIYREKETGQILINHDLATIKEKIPFVRKRGVGVFQDKLLPKLITFENVAF
ALEVICQPSVKKRVLEVLDLVQLKHAKQAFPPDQLSGGEGQQRVSARSIVNNPDVVI
```

```
ADPTGNLDPDTSWEVMKTLSEINNRRGTTVMATHNKEIVNTMKKEVIAIEDGIIVRD
BSRGEYSYD"
3476. .4385
/gene="ftsX"
RBS
3476. .3484
/gene="ftsX"
CDS
3495. .4385
/gene="ftsX"
/codon_start=1
/transl_table=11
/product="cell division protein"
/protein_id="AAC67264.1"
/db_xref="GI:2618835"
/translation="MIKILGRHLRESFKSLGRNTWMTFASISAVTVTLILVGVFLVIM
LNLNWTAEKQVEIKVLDITADQKADQKLDNDIKELKGIQSVFSSKEKELDQIV
DSFGDSGKSLTMDQENLDFVVKITDPDTPNVAKKIEKMDHVKVITYGKEEVSIR
LKVYGVSNIGIALIIGLVFTAMFLISNTIKITIFARRKEBIMKLGVATNFWIRMP
LFLEGILLGVFGSVIPIALVLSITYQYVIGVWVPKVQGSFVSLPYTPNPFVQVSLVLIA
IGAVIGWGLSTSIKKFLRV"
4511. .5967
/gene="yvjB"
RBS
4511. .4516
/gene="yvjB"
CDS
4525. .5967
/gene="yvjB"
/codon_start=1
/transl_table=11
/product="putative protease"
/protein_id="AAC67263.1"
/db_xref="GI:2618834"
/translation="MNQIMAVIAAGSMFLPGGAGVYAGINLLEMDKXPQTAAPATAQA
DSERKMDKIEKAYELISNEYVEKVDREKLLEGALQGMSTLNDPSVYMDKQTAQK
FSDSLDSSEGGIAGVGMEDGKLIIVSPFKSPAEGAKGLKPNDEIISINGESMAGDL
NHAULKIRGKGGSSVMKIQRPCTKKQLSFRIRKABIPLETIVFASKKVQGVHGYIA
ISTFSEHTAEDFAKALRELEKEIEGLVDIVRNGPGYIQSVVEEILKHFVTKDQPYIQ
GTVQOAPMGDGSNKLITLYKWLTPNGNIHKGIPTTAKQPDVFSAGPLQKLEPL
IAERNQKKRYFSTLTKKAYPNVITDKGSASAILAGALKAEAGHYDVVGDTSTFK
KVMNNEDEVKHAQVLLKGLSDPGREDGYFSKDMKAVAFQDNKLNKTGVIDTRTA
ETLNOQIEKSKSDERNDLQQLTAKLSLFVN"
6773. .7977
/gene="yvjD"
RBS
6773. .6777
/gene="yvjD"
CDS
6784. .7977
/gene="yvjD"
/notes="transmembrane protein"
/codon_start=1
/transl_table=11
/product="yvjD"
/protein_id="AAC67265.1"
/db_xref="GI:2618836"
/translation="MSVQNGIELLSAGLFFLHPLFWFFIITLARGYVRIKERTKF
HTRIADYDDLKFTYTKGLIPGLLSVILGGIGISIPGLLAIIVITAAAFTRLAN
WMSAAYIVSVMLIGFLQIYQAEPLERFPQGFVAVPAVAVFLGLLITTEGAVYR
SAHVTSALVSVSSRLPIGQOLANRWLLPLFLVPGNGLESHLWMPVFTVPGGSF
HELIWPFYVFGQGVQGSIPETIRITAKVCVILGAVAVLGASLLWTLPLAGAASCT
ALLGRPLSIKORVNDNAAPFPFSKDDQGLMWLGIIPTNPDAEDLELKGILITKNGI
PVKNVSDFYALQHNKAYVKLEIILNGELRFPQDRASEYEGEHHELGLFLVKKDDREDA
VASGS"
8025. .9373
/gene="yvkA"
complement (8025. .9359)
/gene="yvkA"
/notes="belongs to the major facilitator family"
/codon_start=1
/transl_table=11
/product="yvkA"
/protein_id="AAC67266.1"
/db_xref="GI:2618837"
/translation="MSSRKKEALIVLLGAILVLPINSTMIAVALSSISHTYNESIAS
ITWVTVYIIVMAVQPIAGKGLDMYGNKTMVLMVGLFLIASLGCALSPSLLLIVF
RALQAVGGALPNSIATIRHVVSSEKRLPKVFGFGLGAGLGAALGPFISILIDSFS
WHSIFWNTPFLAIALFTALTMFPQYKENKSDAPLDIIGSLLAGSIVSILITKNEA
```



```
PMGVTYVSVLILLVPLFFRRKRTOHPIDPALFKSSTFTNANLVLNLMMYAVL
LIMPLFNTOGNTSNMGALVSFISFMSASNNVGAOLHKKKAKKIIIFLSPAMAG
ANLIFLILSSSSVLFMLSLILGLASGVGLTSMQVSLATDPGMSGVAGSIFSTF
RYFGSIISALIGLISGYHTLFMLIPAVSIIGVSVLSIGIKSDETARIEKNSA"
RBS complement(9366..9373)
gene /gene="yvkA"
complement(9390..9972)
/genes="yvkB"
/notes="9959"
CDS complement(9390..9959)
/gene="yvkB"
/notes="belongs to the TetR/AcrR family, 5' transcriptional
regulators"
/codon_start=1
Query Match 100.0%; Score 753; DB 1; Length 47739;
Best Local Similarity 100.0%; Pred. No. 5.3e-208;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGAATGAGATGACCGGAATGGTAAACGGAAGAGTCTGTGCAATTTTATGCTGAGGCA 60
Db 23747 GTGAATGAGATGACCGGAATGGTAAACGGAAGAGTCTGTGCAATTTTATGCTGAGGCA 23806
QY 61 TTAACAGAAACTGCAGAGAAATATTTGAACGGCGCAGGCATGTTTTGGTGGGATCAGC 120
Db 23807 TTAACAGAAACTGCAGAGAAATATTTGAACGGCGCAGGCATGTTTTGGTGGGATCAGC 23866
QY 121 CCATTTAACAGCAGTTTTCAGAGGATATATTTACAGATTAATTTGATGGCGGCAAGCT 180
Db 23867 CCATTTAACAGCAGTTTTCAGAGGATATATTTACAGATTAATTTGATGGCGGCAAGCT 23926
QY 181 CAATTTAAAGCGTTTCAGTTTACTTCGAGGCGATGAGCGCGCTAATCTTTAGAACGC 240
Db 23927 CAATTTAAAGCGTTTCAGTTTACTTCGAGGCGATGAGCGCGCTAATCTTTAGAACGC 23986
QY 241 CTTGGAACCTCGAGAGAAAGCTGAACGAAAGTAGGAAGAGGTATCAGGAAACAGG 300
Db 23987 CTTGGAACCTCGAGAGAAAGCTGAACGAAAGTAGGAAGAGGTATCAGGAAACAGG 24046
QY 301 AGATTTTCAGAAAGAGCCCTTGTCGCTCATGCGGGGATCCGAGGCGGATTCATACATTT 360
Db 24047 AGATTTTCAGAAAGAGCCCTTGTCGCTCATGCGGGGATCCGAGGCGGATTCATACATTT 24106
QY 361 TCTGATTTTATAGATAACAAAGCCTACAGCTGTGTGAGACAAAGTAGTGAACATGCAATTT 420
Db 24107 TCTGATTTTATAGATAACAAAGCCTACAGCTGTGTGAGACAAAGTAGTGAACATGCAATTT 24166
QY 421 TTTGAGCAGCCTCATTTTCACATGCTTGTGACATGCTCTGTGAAGCGATAATCGGG 480
Db 24167 TTTGAGCAGCCTCATTTTCACATGCTTGTGACATGCTCTGTGAAGCGATAATCGGG 24226
QY 481 CGTGGCGGGCGCTCAGTTTGATGATGATGAAAGTCAAGTCAAGTATGCTGAAATTTGGCT 540
Db 24227 CGTGGCGGGCGCTCAGTTTGATGATGATGAAAGTCAAGTCAAGTATGCTGAAATTTGGCT 24286
QY 541 GTGGAATATGCTAGTGCAGCTGCCCTTTTATCCGAGCTCCGAGTATTTAGAGGTG 600
Db 24287 GTGGAATATGCTAGTGCAGCTGCCCTTTTATCCGAGCTCCGAGTATTTAGAGGTG 24346
QY 601 GAAGAGACATCCTTTGCTTATCATCTGCTGCGTGAAGCTGGGTGAGAAAGATCAGTAACCAT 660
Db 24347 GAAGAGACATCCTTTGCTTATCATCTGCTGCGTGAAGCTGGGTGAGAAAGATCAGTAACCAT 24406
QY 661 GAATTTTCTATTTGATGCGGCGGAATCAAGGATCTCTATTGTACAGGAATGGCGCAG 720
Db 24407 GAATTTTCTATTTGATGCGGCGGAATCAAGGATCTCTATTGTACAGGAATGGCGCAG 24466
QY 721 ATGCTTTCTGAGAAACGGATCACATCTGAAGGA 753
Db 24467 ATGCTTTCTGAGAAACGGATCACATCTGAAGGA 24499
```

RESULT 2

```
BSUB0018/c
LOCUS
DEFINITION
Accession 194692 bp DNA linear BCT 07-JUL-2003
BSUB0018
Bacillus subtilis complete genome (section 18 of 21): from 3414339
to 3609030.
VERSION 299121 AL009126
299121.2 GI:32468825
SOURCE
ORGANISM
REFERENCE
AUTHORS
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Encian, K.D., Errington, J., Fabret, C., Ferrazi, E.,
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Liu, H.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T.,
Scanlan, E., Schleicher, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serro, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognotti, A.,
Tosato, V., Uchiyama, S., Vandenberg, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzengraber, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and
Danchin, A.
The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
98044033
9384377
2 (bases 1 to 194692)
Direct Submission
AUTHORS
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
REFERENCE
Genetic des Genomes Bacteriens, 28 rue du Docteur Pasteur,
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr.
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
COMMENT
On Jul 7, 2003 this sequence version replaced gi:2635827.
This entry contains data from release R16.1 of the Subtilist
database. Further data on gene annotation and detailed information
about changes from previous releases can be found at
http://genolist.pasteur.fr/Subtilist/.
FEATURES
Location/Qualifiers
1..194692
/organism="Bacillus subtilis subsp. subtilis str. 168"
/mol_type="genomic DNA"
/strain="168"
/db_xref="taxon:224308"
complement(79..888)
/gene="fluC"
/locus_tag="BSU33290"
complement(79..888)
/gene="fluC"
/locus_tag="BSU33290"
/codon_start=1
/transl_table=11
```

```
/product="ferrichrome ABC transporter (ATP-binding protein)"
/protein_id="CAB15335.1"
/db_xref="GI:2635843"
/db_xref="GOA:P49938"
/db_xref="Swiss-Prot:P49938"
/translations="MNSLSEQIGIGYGRVIVEDLNISIPKGIKTLIGPMGCGKST
ILKTMRSIRSHAGAVYLGKAIRKMSDKDIADKMAILPOTPEAPSLTVHELVSYGR
FPHQSGRLNDRILIKWALEETGMAEYERPIEALSGGORQRYWIAALAQGTTEL
LLDEFTYLDLAQLHEILQLDLNKEQRTILMVHDLNHAARFSHYWIALKKGTV
KEGTALVMTDPDLKQVQIDAEIVTDRPNKPVCLTYDLIKNEKELTVV"
complement(904. .1914)
/gene="fhuG"
/locus_tag="BSU33300"
/complement(904. .1914)
/gene="fhuG"
/locus_tag="BSU33300"
/codon_start=1
/transl_table=11
/product="ferrichrome ABC transporter (permease)"
/protein_id="CAB15336.1"
/db_xref="GI:2635844"
/db_xref="GOA:P49937"
/db_xref="Swiss-Prot:P49937"
/translations="MKTNKPILVMAVFLLIWVVFISLNLGVIKIAPLKTQVFF
GQCTARDELVFRFLPRILISLLVAGISVAGAILQSVQNELAEPGILINAGGSL
AVVLYFYFFGSGSADSLSPFGFTMLPFSALAGAILAFLIYLAWKKGVTPIRLIYGI
GWAGNALLILLIQKMDPHDPMQAAWISGSINGANWMMIALLPWIILLITLYIK
ARYLINQDQGLATGLTGAKEKERRILLAAVTLAASCVAAGGIAFLGLIAPHVAR
RLTGPRHQTILVPSAFISGSLFLADTLARNVLAPSEIPVGLVISLGAIFYIYLLMK
AN"
complement(1914. .3068)
/gene="fhuB"
/locus_tag="BSU33310"
/complement(1914. .3068)
/gene="fhuB"
/locus_tag="BSU33310"
/codon_start=1
/transl_table=11
/product="ferrichrome ABC transporter (permease)"
/protein_id="CAB15337.1"
/db_xref="GI:2635845"
/db_xref="GOA:P49936"
/db_xref="Swiss-Prot:P49936"
/translations="MHFHCSSKHSIKSAEKSDILKQOLIIISNRKEVRQLSOKNIR
TASEEIQWTSRTYGAIVLVLAGICLGLAFLSISGAADIHURTVWEAIFHYQPTKT
SHQI IHDRLRPTAAALVAGLAVGAINMQGTRNPLAEPSTMGVTSFAVSIAP
APFPGLSAMGLVLSFAGLGAFTVMGIGMFRGGITPVKLALAGTAVTYPFTGIST
AIAIRFDVAODIGFWYAGGVAGVKWGVQLLIAGAVGLTAPFIARSVTVLSLGDLL
AKLGOYTSAVKLGVMLIVVLGAAVSIAGTIAFLIIPHITRFLVGVYXWIIIPC
SAVLGAVLLVFADIAARLVNAPETPVGALTSLIGVPPFFYLARRRRL"
3166. .4113
/gene="fhuD"
/locus_tag="BSU33320"
/3166. .4113
/gene="fhuD"
/locus_tag="BSU33320"
/codon_start=1
/transl_table=11
/product="ferrichrome ABC transporter (ferrichrome-binding protein)"
/protein_id="CAB15338.1"
/db_xref="GI:2635846"
/db_xref="GOA:P37580"
/db_xref="Swiss-Prot:P37580"
/translations="MTHIYKKGAFPAFLILLIILAACGNNSBSKGSASDSKGAETFT
YKAGTSVKIPKPKRVVWADGYVFKTLGINVVGAPENVKFNPKYKNGKTVGNI
DGCTSVKVDLNPDLIIVNTQGADIKLEKIAPTVAVKYKLDNILOLEKFAKWTG
TEDNAEKWLAKDKKVAARAKTKKAVGDKIISIMQINGKDIYVFGKDFRGGSIIYK
DLGLQATPKLKEKADGGPGYTSISLEKLPDFAGDIYFAPGWQSGDDGGVFPSSIWK
NLNAVKNHVKYKMDPIGYFTDPISLEQLEFITESLTK"
complement(4348. .5757)
/gene="yvhH"
```

```
/locus_tag="BSU33330"
complement(4348. .5757)
/gene="yvhH"
/locus_tag="BSU33330"
/function="similar to ABC transporter (amino acid permease)"
/notes="similar to ABC transporter (amino acid permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB15339.1"
/db_xref="GI:2635847"
/db_xref="GOA:O32204"
/db_xref="TrEMBL:O32204"
/translations="MEQTKWGFLLITAFVVGVMVGSIFSLPSLSIASIAPFGATSA
WLTGTAGVLMIALVFGHLSIRKPELTAGPSVARALFSDPKKGNAAAGFTVMWGYWVAS
WISNVAITSLAGYLTSPFPIILVDKREMESISGQEVTLGQLLTFVCTILLWTHAIL
VASINGASKLNFVTTLSKVLGVFFVIFVAGLVFVQTSFLGFHFPYVOGENGTSIGIGSQ
VHNAALSTLWAFVGHIESAVILSGRSQDRVKATITGLLIALSIITVLIITWGVLP
HDKLVGSEKPFVDVLYAVIGNAGSVIMALLAILCLFGMLGILLGSEVYPQAQADP
PPAFTAKTNKGSPIVLIITNMQSVFIFSVISRTISDAFTLTAAITLAIYLIPLYV
SAIYSLKVVIKGETYDQLKGSVRDGLIAILACAYSVFVITGTADLTITFILGILFF
VGLIYVYPVSNKFKQEKQA"
complement(6157. .6297)
/gene="sspJ"
/locus_tag="BSU33340"
/complement(6157. .6297)
/gene="sspJ"
/locus_tag="BSU33340"
/codon_start=1
/transl_table=11
/product="small acid-soluble spore protein"
/protein_id="CAB01467.1"
/db_xref="GI:32468826"
/translations="MGFFNKDKGRSEKKNVIQGADEAGSALKDDPLQEAQKKKN
NR"
6464. .6946
/gene="yvsG"
/locus_tag="BSU33350"
/6464. .6946
/gene="yvsG"
/locus_tag="BSU33350"
/function="similar to unknown proteins"
/notes="similar to unknown proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB15340.1"
/db_xref="GI:2635848"
/db_xref="GOA:O32205"
/db_xref="Swiss-Prot:O32205"
/translations="MTGKTHIMGGIASCTAAAYYYGFDPVLMASGAVGALIPDICT
QSKIGRKFPILSKVSVSFGHRTFTHSLFLMLIMFITSTYIPDKNISAGLMIGMASH
LILDATVNGIKLLFPSTIRVRLPLYMKTGSFSEQLVLAGLTLASCYFFYFMYLFGHMF"
7046. .8899
/gene="yvgJ"
/locus_tag="BSU33360"
/7046. .8899
/gene="yvgJ"
/locus_tag="BSU33360"
```

Query Match	100.0%	Score	753	DB 1	Length	194692	
Best Local Similarity	100.0%	Pred. No.	6.7e-208	Mismatches	0	Gaps	0
Matches	753	Conservative	0	Indels	0	Gaps	0
Qy	1	GTGAATGAGATGACCGGAATGGTAAACGGAAGAAGGTCTCTGTCATTTTATTGCTGAGGCA	60				
Db	189268	GTGAATGAGATGACCGGAATGGTAAACGGAAGAAGGTCTCTGTCATTTTATTGCTGAGGCA	189209				
Qy	61	TTAACAGAAAACCTGCAGAGAAATATTGAAACGGCGCAGGATGTTTTGTTGGTGGGATCAGC	120				
Db	189208	TTAACAGAAAACCTGCAGAGAAATATTGAAACGGCGCAGGATGTTTTGTTGGTGGGATCAGC	189149				
Qy	121	CCATTTAACAGCAGGTTTTTCAGAGGATTATATTACAGATTAAATTCAGATGGCGGAAAGCT	180				
Db	189148	CCATTTAACAGCAGGTTTTTCAGAGGATTATATTACAGATTAAATTCAGATGGCGGAAAGCT	189089				
Qy	181	CAATTTAAAAGCGTTTTTCAGTTTTTACTTTCAGGGGATGAGCGGCGTAATCTTCTAGAACGC	240				

Db 189088 CAATTTAAAGCGTTTCAGTTTCTTTCGAGGCGATAGCGCGTAACTCTTAGAGCG 189029
Qy 241 CTTTGAAGCTCCGAGAGAAAGGCTGAACGAAAGTAAAGAAAGAGGTATCACGAAACAGG 300
Db 189028 CTTTGAAGCTCCGAGAGAAAGGCTGAACGAAAGTAAAGAAAGAGGTATCACGAAACAGG 188969
Qy 301 AGATTTCAGAAAGAGCCCTTGTGGCTCATGCGGGGATCCGAAAGGCGATTCATACATTT 360
Db 188968 AGATTTCAGAAAGAGCCCTTGTGGCTCATGCGGGGATCCGAAAGGCGATTCATACATTT 188909
Qy 361 TCTGATTTATAGATAACAAAGCCTACAGCTGTTGAGACAAGAGTTGAACATGCATTT 420
Db 189008 TCTGATTTATAGATAACAAAGCCTACAGCTGTTGAGACAAGAGTTGAACATGCATTT 188849
Qy 421 TTTGAGCAGCCTCATTTTGCACATGCTGTTTGGACATGCTCTGCTGAAGCGATAATCGGG 480
Db 188848 TTTGAGCAGCCTCATTTTGCACATGCTGTTTGGACATGCTCTGCTGAAGCGATAATCGGG 188789
Qy 481 CGTGGCGGGCGCTCAGTTTGATGATGAAGAGAGTCACTGAGGATATGCTGAATTTGGCT 540
Db 188788 CGTGGCGGGCGCTCAGTTTGATGATGAAGAGAGTCACTGAGGATATGCTGAATTTGGCT 188729
Qy 541 GTGGATATCTCATGCTAGCTGAGCTGCCGTTTTTTATCGAGCTCCGGATATTTAGAGGTG 600
Db 188728 GTGGATATCTCATGCTAGCTGAGCTGCCGTTTTTTATCGAGCTCCGGATATTTAGAGGTG 188669
Qy 601 GAAGAGACACTCCTTGTCTATCATCTGCTGGAAGCTGGGTGAGAGATCAGTAACCAT 660
Db 188668 GAAGAGACACTCCTTGTCTATCATCTGCTGGAAGCTGGGTGAGAGATCAGTAACCAT 188609
Qy 661 GAATTTCTATTTGATGCGGCGGAATCAAGGATATCTATTGTACAGGAAATGGCGCAG 720
Db 188608 GAATTTCTATTTGATGCGGCGGAATCAAGGATATCTATTGTACAGGAAATGGCGCAG 188549
Qy 721 ATGCTTCTCAGAAACGATCAGATCTGAAGGA 753
Db 188548 ATGCTTCTCAGAAACGATCAGATCTGAAGGA 188516

RESULT 3
LOCUS BX571859/c 349483 bp DNA linear BCT 26-SEP-2003
DEFINITION Photorhabdus luminescens subsp. laumondii TT01 complete genome;
segment 1/17.
ACCESSION BX571859 BX470251
VERSION BX571859.1 GI:36783455
KEYWORDS complete genome.
SOURCE Photorhabdus luminescens subsp. laumondii TT01
ORGANISM Photorhabdus luminescens subsp. laumondii TT01
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
REFERENCE 1
AUTHORS Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
Bocs,S., Boursaux-Eude,C., Chandler,M., Dassa,E., Derose,R.,
Derzelle,S., Freydisse,S., Gaudriault,S., Givaudan,A., Glaser,P.,
Medigue,C., Lanois,A., Powell,K., Sigulier,P., Wingate,V.,
Zouine,M., Boenare,N., Danchin,A. and Kunst,F.
Complete genome sequence of the entomopathogenic bacterium
Photorhabdus luminescens
Nat. Biotechnol. 11 (1) (2003) In press
2
AUTHORS Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
Direct Submission
TITLE Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
lfrangeu@pasteur.fr, fkunst@pasteur.fr
LOCATION/Qualifiers
FEATURES
source 1. .349483
/organism="Photorhabdus luminescens subsp. laumondii TT01"
/mol_type="genomic DNA"
/strain="TT01"
/db_xref="taxon:243265"

gene
RBS
CDS
223. .1622
/gene="dnaA"
/locus_tag="plu0001"
223. .228
/gene="dnaA"
/locus_tag="plu0001"
234. .1622
/gene="dnaA"
/locus_tag="plu0001"
/codon_start=1
/transl_table=11
/product="Chromosomal replication initiator protein DnaA"
/protein_id="CAE12296.1"
/db_xref="GI:36783456"
/translation="MSLSLWQCCARLQDELPAEFMSWIRPQLAELSGNTLALYAPN
RFVLDWVRKYINNINGLDFCGAEVPLRFVEVGNKPSQNDSPQPVVHTPTVAPAY
PONTSVPSMDNTAVQPELSYRNVNPKHTFDNVEGKSNQLAARQVADNPGGAY
NPLFLVGGTGLGKTHLLHAGNSIMARKANAKVVMHSEFVODMVKALONNAIEEPK
RYRSFWDALLIDDIQFFANKERSQBEFFHTFALLEGNQOIILTSRYPKEINGVEGR
LKSRFGWGTVAIEPPELETRVALMKADENEIQLPGEVAFIAKRLSRNVELEGA
LNRVIANNTGTRAITIDFVREALRDLALQEKLVTDNIQKTVAEYIKTKVADLLSK
RRSRVARPROMAMALAKELTNHSLPEIGDAFGGRDHTTVLHACRKIEQLREESHDIK
EDFSNLIRTLSS"
1627. .2727
/gene="dnaA"
/locus_tag="plu0002"
1627. .2727
/gene="dnaA"
/locus_tag="plu0002"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, beta chain"
/protein_id="CAE12297.1"
/db_xref="GI:36783457"
/translation="MKFIITEREQLLPQVSSPLSGRPTLPILGNLLLOVTEGSLLL
TGTDLFEMENMARVTLSLPHEVGATVPARKFDFIWLGLPDGAEISVELDGRLLVRSG
RSRSLSTLPASDFPNLDWQSEVSELSQATLKRLEISTQFSMAHQDVRYINGMLF
ETEGEELRTVATDGHRLAVCSMGIGQKPSHVSIVPRKGVIELMRLDDGDTPLQLQI
GNNIRAHVQDFTFSLKVDGRFPDYRVLKPNPKDTLEANCMDLKAQFRAAILNSE
KFRGRLVYFSENLRIANNPEOEAEIIVDVSYQGAEMEIGFNVSVLVDLNTLRCE
KVRLLLTDAISSVQIEDCANHNAAYVVMPEML"
2749. .3840
/gene="recF"
/locus_tag="plu0003"
2749. .3840
/gene="recF"
/locus_tag="plu0003"
/codon_start=1
/transl_table=11
/product="DNA replication and repair protein"
/protein_id="CAE12298.1"
/db_xref="GI:36783458"
/translation="MTLRLIRDFENIAADLPLATGFNVLVNSGSGKTSVLEATY
TIGHGSRFSIOAGRVILHGCDFFVLHGRLOQENERSISIGLSKRNNGSKVRIDGS
DGKIALKAMLPQMLITPEGFTLLNGPKYRFAIDWGCFFHNEPFFSAWNLKRL
KORNAALRVTRYQIRPDWDELIPLANQINRWRYVNTIQDITNTKQFLPEFTL
SFSFQGWKDSYAEELRQEDRDTLTATSGPHKADLRIRAEQTPVEDMLSRGOL
KLIMCALRLAQEYFTQSGOQCLYLLDDFASLDTSRRLQALARKLSTQAVFVSAI
NPQDITDMLDGNKMFREVENKIEVQPD"
3860. .6326
/gene="gyrB"
/locus_tag="plu0004"
3860. .6274
/gene="gyrB"
/locus_tag="plu0004"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="CAE12299.1"
/db_xref="GI:36783459"
/translation="MSNTYDSSSIKVLKGLDAVRKPGMYIGDIDGDTGLHVMFEV
DNAIDBALAGHCSEVITVTHADNSVQDDGRGIPGTIGHEEGVSAAEVNTVLHAGG
KFDNDSYKVSGLGHVGVSVNALSSEKLELVIRRDGKVHEQTVHLGVFQSPKLVVGET

```

EOTGTRVRFPWSPMDTFRDNTPEFQHDILAKELRELFLNSGVSRILDKRTNIBDHPHY
EGGKAFVEFLNKNKTPIHENVFYFTEKDGIGVEISOMWNGDFQENIYCFNTNIPOR
DGGTHLVFRTATNRTLNSYMDKEGYNKKSVSATGDDAREGLAVIVSVKVPDPKFS
QTKDLVSSEVKTAVETLMMKLVLYLLENPNDAKTUVSKIIDAAAREARAKAREMT
RRKGALDLAGLPKGLQCDQDRPALBELYLVEGDSAGGSAKQGRNRKNOAIDPLKGI
LNVEKARFDMKLSQSEVATLITLALGCGIGRDEYNPKLRYHSIIMTDADVDSHIRT
LLITFFYRWPEPIIERGHVFIAPPLYKVKKGQEQVIKODEAMDEYQLSIALDGASL
YTNHVPALRGEPLKLVETFNQVOKIIRMERLYPLSLNLSLIYHPNLTEETLSDKA
KVEWEMLVYLRDLNLESENGSSYSYBLHENRERQVPEVLCVPTHGVDYDILLDFDV
HGGEYHRITLLGDKLRNLIEEDAYIERGERRQVPSFEQALNWLTKESRGLUSIQRYK
GLGEMPEQWETTMNTPETRMVRVTYKDAIATDQLFTLLMGDAVEPRRAFIEENALK
AANIDI".6326
/terminator
/genes="gyrB"
/locus_tag="plu0004"
6415..6926
/locus_tag="plu0005"
6415..6894
/locus_tag="plu0005"
/function="Unknown"
/notes="unnamed protein product; Similar to Hcp protein"
/codon_start=1
/transl_table=11
/protein_id="CAE12300.1"
/db_xref="GI:36783460"
/translations="MADMIYMTIKGKKOGLISAGCSSVDSIGNKYQANHFDOILVYSL
SHAIRTQNVHDQPIILQIKFDKSSPLLGVAISENSLSCSIDPYETNSNGAQRVYN
IKITGARITISIVLHPSLHNHLELPQESISLKYNTITWNHLIAGTGGYSIMDDRYY"
6908..6926
/locus_tag="plu0005"
/complement(8013..8778)
/locus_tag="plu0006"
/complement(8013..8768)
/locus_tag="plu0006"
/function="Unknown"
/notes="unnamed protein product; Similar to unknown
protein"
/codon_start=1
/transl_table=11
/protein_id="CAE12301.1"
/db_xref="GI:36783461"
/translations="MKQDNLLDLKVRRELAKTLSPVLTHSKIRDMFNSKINLFVE
DILOQVRIHCLSGNATIEIIOFEDHLKEQAPCLTANRVKQYAIIEREKESYTNLI
LKQIFVGGGTQILAGYVCKASIGLACASFGAPLMAHGYNVNVGVYLLYRENING
GVREGRYVANKGLSDXADITATVDLVLSGIGIFRVLKPREKSNLSFRNINSDF
IRWRENGVSLTTEGIVDVSTLSIYQLEKEK"
/complement(8773..8778)
/locus_tag="plu0006"
/complement(8935..10056)
/locus_tag="plu0007"
/complement(8935..10044)
/genes="asd"
/locus_tag="plu0007"
/locus_tag="plu0007"
/codon_start=1
/transl_table=11
/product="aspartate-semialdehyde dehydrogenase (ASA
dehydrogenase) (ASADH)"
/protein_id="CAE12302.1"
/db_xref="GI:36783462"
/translations="MKKNVGFIGWRGMVGSVMQRMIEEDQFDVICPFFPTSHQGP
APDFTGQGLQNAFDIEALGALDIIISCOGGDYTNEIYPKLRATGQGYWIDAASAL
RMNDSIILDPNHTHIQQGLNGIKTFVGGNCTYSMLMSLGLPNDLVWASVA
TYQASGAGHRELLVQMSLHTQVAKELQDPASAILDIERKVDFTKSGVNSTEK
FSVPLAGSLIPWIDKQDNGOSREEMKGOAETNKILNTGNIIITVDGLCVRI GALRCH
SOAFTLKLKDIISIPETEQLLAAHNDWVRVIPNDRELMSRELTPAAVTGLDTPVGLR
KLNMGPEVLSAFTVQQLLWGAELPLRLMLRILI"
/complement(10051..10056)
/genes="asd"
/locus_tag="plu0007"
/complement(10251..10472)
/genes="ogrK"
/locus_tag="plu0008"

```

```

complement(10251..10472)
/genes="ogrK"
/locus_tag="plu0008"
/function="Unknown, probable positive regulator of phage
P2 late gene transcription"
/codon_start=1
/transl_table=11
/product="Ogr protein"
/protein_id="CAE12303.1"
/db_xref="GI:36783463"
/translations="MIKPLCCGAHAHARSFHSCHTKERYNQCQININGATFVSHET
FVRFTSKPGEVISVVKPEKTKTQLNVT"
/complement(10549..11634)
/locus_tag="plu0009"
/complement(10549..11634)
/locus_tag="plu0009"
/function="Unknown, probable tail synthesis"
/notes="unnamed protein product; Similar to protein D (gpd)
of Enterobacteria phage P2"
/codon_start=1
/transl_table=11
/protein_id="CAE12304.1"
/db_xref="GI:36783464"
Query Match 8.2%; Score 61.6; DB 1; Length 349483;
Best Local Similarity 46.4%; Pred. No. 3.9e-06;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;
CDS
QY 46 TTTATTGCTGAGGCATTAACAGAAAACCTGCAGAGAAATATTTGAACGGCGCAGGCATGTT 105
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315652 TTTACTGTCCAAAGGTGAAACCTCTCGTTGTGTGACCAATATTATCAAAAAGGTGATCATCGCG 315593
QY 106 TTGGTGGGATCAGCCCATTTAACAGCAGGTTTTACAGAGATTTATTTACAGATTAATT 165
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315592 CTAATAGGATAAGCCCTTTAACTCGCGTTTTTTTCAAAAAGACTATGTAGTGGACCTTATT 315533
QY 166 GGATGGGCGAAAGCTCAATT---TAAAGCGTTTTCAGTTTACTTGCAGGGCATGAGGCG 222
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315532 CAGTGTCAAGTCATTTATTTCCACAGTCGACATATTATCTTGTGAACGTGAAGCT 315473
QY 223 GCTAATCTTCTAGAACGGCTTGGAACCTCCGAGAGGAAGACTGAAACGAAAAGTAAGAAA 282
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315472 TCACGCCCTTTAGTCGTAGTGAATTGATAATGTTTAAAGCTATCAAAAACACATCGC 315413
QY 283 GAGGTATACGAAACAGGAGATTTCAGAAAGAGCCCTGTGGC-----TCATGGCGGGG 337
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315412 GAAATTAGACGTCAITTTACGTAACTTGATTTATGTTATTTCCACAGCAACATTTGAAAAGT 315353
QY 338 ATCCGAAAGCGATTTCATAC-ATTTTCTGATTTTATAGATAACAAAGCTACCAGCTGTTG 396
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315352 AGCAATCAGAGTCATCCAAATTTAGTGACTTTTCACTAAACCATGACTACCAATCTCT 315293
QY 397 AGACAAAGAGTTGAAACATGCATTTTGTGACAGCCCTCATTTTCGACATGCTTGTGGAC 456
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315292 AAAACACAAGTTGAAAACCGGTTTAATGAATCAGAAATCTTTTAAAAAAGCTGTCTTGAT 315233
QY 457 ATGCTCTGTGAGCGATAATCGGGCGTCGCGGGCGCTCAGTTTGTATGATGGAAGATC 516
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315232 ATGCTCTTCAAGCCATAAAGGGCGCACTAAAGAGTACTGGGCAATACCTTTTGTGCAAA 315173
QY 517 AGTGAGGATATGCTGAATTTGGCTGTGGAAATATGTCTAGCTGAGCTGCCGCTTTTTTATC 576
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315172 GACCTACAATTAGTATATAAAGGTTGCCATATATTTCGCTGAAATCTCTTTTACCTC 315113
QY 577 GGAGCTCCGGATATTTTAGAGGTGGAAGAGACACTCTCTGCTTATCATTCGTCGTGGAAG 636
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315112 AATACCCCTCGATTACTTGGGGTAAAGTATCTTACGTACTTATACCCCTCTGGTCA 315053
QY 637 CTGGGTGAGAAGATCATGACCAATGAATTTTCTATTGATGCGCGCGATCAAGGATAT 696
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315052 ATCGAAAAGGGTATTTTAACGGTAGTTATCTCTATACAAGTAGCAGATAAACAAGTTAC 314993
QY 697 CTCATTGT 704

```

```
Db 314992 GGAATCGT 314985 |||||
RESULT 4
AX770907
LOCUS AX770907 349980 bp DNA linear PAT 02-JUL-2003
DEFINITION Sequence 38 from Patent WO02094867.
ACCESSION AX770907
VERSION AX770907.1 GI:32438071
KEYWORDS
SOURCE
ORGANISM
Photorhabdus luminescens
Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
REFERENCE
1
AUTHORS Duchaud,E., Taourit,S., Glaser,P., Frangeul,L., Kunst,F.,
Danchin,A. and Buchrieser,C.
TITLE Sequence of the Photorhabdus luminescens strain TT01 genome and
uses
JOURNAL Patent: WO 02094867-A 38 28-NOV-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
Location/Qualifiers
1. 349980
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/db_xref="taxon:29488"
/notes="sequence length to big . Splitted into 2
sequences.--seq 00038: 1 too 349980--seq 10787: 300001 to
618776"
ORIGIN
Query Match 8.2%; Score 61.6; DB 6; Length 349980;
Best Local Similarity 46.4%; Pred. No. 3.9e-06;
Matches 310; Conservative 0; Mismatches 349; Indels 9; Gaps 3;
Qy 46 TTTATTGCTGAGGCAATTAACAGAAACTGCAGAGAAATATTGAAACGGCGCAGCGCATGTT 105
Db 282443 TTTACTGCTCAAGGTGAACCTCTCGTTGTGACCAATTTTCAAAAGGTGATCACGCG 282502
Qy 106 TTGGTGGGATCAGCCATTAACAGCAGGTTTTCAGAGATATATTACAGATTAATT 165
Db 282503 CTAATAGGAGTAAGCCCTTTAACTCGCGTTTTTCAAAAGCACTATGTAGTGGACCTTATT 282562
Qy 166 GGATGGCGGAAAGCTCAATT---TAAAGCGTTTCAGTTTACTTTCAGGCGCATGAGCGG 222
Db 282563 CAGTGGTCAAGTCATNTATTTCCGCAAGTGCACATATATTATACCTTGTGAACGTGAAGCT 282622
Qy 223 GCTAATCTCTAGAACGCTTGGAACTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAA 282
Db 282623 TCAGCGCTTTTAGTCGCTAGTGGAAATGATAATGTTAAAGCTATCAAAAACACATCGC 282682
Qy 283 GAGTATCAGAAACAGGAGATTTGCAAGAGAGCCCTTGTGGC-----TCATGGCGGGG 337
Db 282683 GAAATTAGACGCTATTACGTAACTTGATTATGTTATTTCCACAGCAACATTGAAAGT 282742
Qy 338 ATCGRAGGCGATTATAC-ATTTCTGATTTTATAGATAACAAGCTACCGACTGTG 396
Db 282743 AGCAAAATCAGATCATCCAAATTTAGTGACTTTTCACTAAACCATGACTACCAATCTCTT 282802
Qy 397 AGACAAGAAGTTGAACATGCAATTTTTTGGAGCAGCTCATTTTCACATGCTTGTGTTGGAC 456
Db 282803 AAAACACAAGTTGAAACGGCTTTAATGAATCATCAATCTTTTAAAAAAGCTGCTTGAT 282862
Qy 457 ATGTCTCGTGAAGCGATAATCGGCGGTGCGCGGGCGTCAAGTTTGATGAGGAAGATC 516
Db 282863 ATGTCTTTCAGGCCATAAAGGGCGACTAAAGGTAAGGTAAGTACTGGCAATACCTTTGGTCAAAT 282922
Qy 517 AGTGAGGATATGCTGAATTTGGCTGTGGAATATGTCATAGTGCAGCTGCCGTTTTTATC 576
Db 282923 GACCTACAATTAGTATATAAGCGGTGCCATATATTTTCGTCGAATTTCTCTTTTACCTC 282982

Qy 577 GGAGCTCCGGATATTTTAGAGGTGGAAGACACTCTTCTGCTTATCATCTGCTCGTGAAG 636
Db 282983 AATACCCCTCGATTACTTGGGTAAGATTTCTACGTACTTATCATCCGCCCTTGTCA 283042
Qy 637 CTGGGTGAGAAGATCAGTAACCATGAATTTTCTATTGTTATCGGCGCGGAATCAAGGTAT 696
Db 283043 ATCGGAAAAGGTTATTTAACGGTAGTTATCTTATACAGTACGAGATAACAAGTTAC 283102
Qy 697 CTCATTGT 704
Db 283103 GGAATCGT 283110

RESULT 5
I66494/c
LOCUS I66494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 7218)
AUTHORS Dörner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
Location/Qualifiers
1. 7218
source
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 6.5%; Score 48.6; DB 6; Length 7218;
Best Local Similarity 4.3%; Pred. No. 0.012;
Matches 15; Conservative 195; Mismatches 139; Indels 0; Gaps 0;
Qy 1 GTCAATGAGATGACCGGAATGGTAACGAAAGAGTCTGTGCATTTTATGCTGAGGCA 60
Db 1460 GTTAAAGAGATAGAAGAAATTTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1401
Qy 61 TTAACAGAAACTGCAGAGAAATATTGTAACGGCGCAGGCATGTTTGTGGGATCAGC 120
Db 1400 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1341
Qy 121 CCATTAAACAGCAGGTTTTCAGAGGATATATTACAGATTAATTGATCGCGGAAAGCT 180
Db 1340 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1281
Qy 181 CAATTTAAAGCGCTTTCAGTTTTCAGGCGCATGAGCGGCTAATCTTCTAGAAGCG 240
Db 1280 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1221
Qy 241 CTTGGAACCTCCGAGAGAAAGGCTGAACGAAAGTAAGGAAAGAGGTATCAGCAACAGG 300
Db 1220 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1161
Qy 301 AGATTTCAGAAAGAGCCCTTGTGGCTCATGGGGGATCGGAGGCGA 349
Db 1160 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112

RESULT 6
AC040918/c
LOCUS AC040918 175919 bp DNA linear PRI 16-FEB-2002
DEFINITION Homo sapiens chromosome 15, clone RP11-684B21, complete sequence.
ACCESSION AC040918
VERSION AC040918.7 GI:18693519
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```



```
10381325
PUBMED
REMARK Submitted to Molecular Phylogenetics and Evolution
REFERENCE 2 (bases 1 to 1041)
AUTHORS Omland,K.E., Lanyon,S.M. and Fritz,S.J.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1998) Bell Museum of Natural History, University
of Minnesota, 1987 Upper Buford Circle, Saint Paul, Minnesota
55108, USA
FEATURES
source Location/Qualifiers
1..1041
/organism="Amblycercus holosericeus"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:84791"
1..1041
/notes="NADH-2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AA03308.1"
/db_xref="GI:6090722"
/translation="MPRASLIPTTSLLGTTTITSSNHWMAWTGLEINTLAILPLI
SKSHPRALAEAAKYLTOAASALVLFSSMTNMTGQWDITQLTHPTSCLLTSAI
AMKLGVPFHFPEVLQGSPLTGLLSTIMKLPITLLYMTSPSLNPLTLTIL
STALCGWMLNQTQIRKILAFSSISHLGMWAIILVNPKLITILFYLAMMTATIELT
LNTIKVLKSLTMTAKTSPSLNMLLTLTSLAGLPPLTGPLPKLILIOELTKQMA
PAATLIALLSLLSPFLRLAYCATITLPHTTNHWKQRTNKPNTITAILITVSAI
LLPISPMILTIF"
ORIGIN
Query Match 5.7%; Score 42.8; DB 5; Length 1041;
Best Local Similarity 54.4%; Pred. No. 0.44; Indels 0; Gaps 0;
Matches 86; Conservative 0; Mismatches 72;
Qy 585 GGAATATTTAGAGTGAAGACACTCTTCCTTATCATCGTCGCGGAAGCTGGGTGA 644
Db 603 GGTGAGTTTGGGTTGTAGATGATAATGATGCTATCCTCCTAGGTGGGAGATGA 544
Qy 645 GAGATCAGTAACGATTTCTATTGTATGCGGCCGAATCAAGGGTATCTCATGT 704
Db 543 AAAGGCTAGGATTTTCGGATTTGTGTTGTTGAGGCCCAATCAACCTCCGAGGGCTGT 484
Qy 705 ACAGGAATCGCGCAGATCTTTCTGAGAAACGGATCA 742
Db 483 AGAGAGATCGCTAAGTGTTTAATAGTGTGGGTTCA 446
RESULT 9
E08319
LOCUS E08319 1659 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding luciferase from firefly.
ACCESSION E08319
VERSION E08319.1 GI:2176437
KEYWORDS JP 1994030982-A/1.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1659)
AUTHORS Yoshino,S., Shiraishi,S., Inoue,S. and Saigo,K.
TITLE FIREFLY ENZYME LUCIFERASE GENE
JOURNAL Patent: JP 1994030982-A 1 01-NOV-1994;
CHISSO CORP
COMMENT OS Photuris (firefly)
PN JP 1994030982-A/1
PD 01-NOV-1994
PF 21-APR-1993 JP 1993119050
PI YOSHINO SHUHEI, SHIRAISHI SHINJI, INOUE SATOSHI, SAIGO KAORU
PC C12N15/53,C12N1/21,C12N9/02,C12N15/70,(C12N1/21,C12R1.19), PC
(C12N9/02,
PC C12R1.19);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
```

```
CC anti-sense: No; Location/Qualifiers
FH Key
FT source 1..1659
/organisms="Photuris"
/clone="ppFL7"
FT CDS
FT 1..1659
/product="Luciferase from firefly".
FEATURES
source Location/Qualifiers
1..1659
/organism="unidentified"
/mol_type="genomic RNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 5.6%; Score 41.8; DB 6; Length 1659;
Best Local Similarity 29.3%; Pred. No. 0.94;
Matches 90; Conservative 33; Mismatches 184; Indels 0; Gaps 0;
Qy 307 GCAGAAAGAGCCCTTGCTCATGCGGGGATCCGAGCGGATTCATACATTTCTGAT 366
Db 1258 GNGAYATGGNTAYTTYGARGAYGNCAYGTTATATGTCNGAYNGNNTNARNNN 1317
Qy 367 TTTATAGATAACAAGCCCTACCGCTGTTTGAGACAAAGTGTGAACATGCAATTTT 426
Db 1318 NTNATHAARTAYAARGNTAYCARGTNCNCNGCNGARNTNGARGCNCNTNNTN 1377
Qy 427 CAGCCTCATTTGCGACATGCTTGTGTCGACATGCTCTCGTGAACGGATAATCGG 486
Db 1378 CAYCCTTAYTGARGAYGNGGNGTNGCNGGTCNCNGAYGARGTNGCNGNGAY 1437
Qy 487 CGGGGCGTCAGTTTGATGATGGAAGATCAGTGAGGATATGCTGAATTTGGCTG 546
Db 1438 CNGGNGCNGTNGTNGTNTNNAARGGNAARNNNATHACNGARARATHCARGAY 1497
Qy 547 TATGTCATAGCTGAGCTGCGCTTTTATCGGAGCTCCGGATATTTAGAGGTGGAAG 606
Db 1498 TAYGTNGCNGCARGCTNACNNNNNNAARAENTNNGGNGGNGTNGARTTYGTNA 1557
Qy 607 ACACTCC 613
Db 1558 GARGTNC 1564
RESULT 10
AF526473/c
LOCUS AF526473 1006 bp DNA linear VRT 12-NOV-2002
DEFINITION Motacilla flaviiventris specimen-voucher FMNH352834 NADH
dehydrogenase subunit 2 (ND2) gene, partial cds; mitochondrial gene
for mitochondrial product.
ACCESSION AF526473
VERSION AF526473.1 GI:24899324
KEYWORDS
SOURCE
ORGANISM mitochondrion Motacilla flaviiventris (Madagascar wagtail)
Motacilla flaviiventris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Passeriformes; Motacillidae;
Motacilla.
REFERENCE 1 (bases 1 to 1006)
AUTHORS Voelker,G.
TITLE Systematics and Historical Biogeography of Wagtails: Dispersal
Versus Vicariance Revisited
JOURNAL Condor 104 (4), 725-739 (2002)
REFERENCE 2 (bases 1 to 1006)
AUTHORS Voelker,G.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2002) Barrick Museum of Natural History,
University of Nevada Las Vegas, 4505 Maryland Parkway, Las Vegas,
NV 89154, USA
FEATURES
source Location/Qualifiers
1..1006
/organism="Motacilla flaviiventris"
/organelle="mitochondrion"
```



```
Db 1495 GAYTAYTNGCNGCARGTACGTCNACNTCTCNAAARAATYTNMGNGCGNGTGNGARTTYGTN 1554
Qy 604 GAGACACTCC 613
Db 1555 AARGARGTNC 1564

RESULT 13
LOCUS AY294513/c 1056 bp DNA linear PAT 13-MAY-1997
DEFINITION Aerodramus terraereginae specimen-voucher DHC20 NADH dehydrogenase
ACCESSION AY294513
VERSION AY294513.1 GI:2083170
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1659)
AUTHORS Zenko,S., Shiraishi,S., Inouye,S. and Saigo,K.
TITLE Photuris firefly luciferase gene
JOURNAL Patent: US 5618722-A 2 08-APR-1997;
FEATURES
source Location/Qualifiers
1..1659
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 5.5%; Score 41.2; DB 6; Length 1659;
Best Local Similarity 30.0%; Pred. No. 1.4;
Matches 93; Conservative 39; Mismatches 178; Indels 0; Gaps 0;
Qy 304 TTTCGAGAAAGAGCCCTTGGCTCATGCGGGGATCCGAGGGGATTCAATTTCT 363
Db 1255 TCNGGNGAYATHGGNTAYTTYGAYGARGAYGNCAYGNTAYATHGNGAYMGNYTNAAR 1314
Qy 364 GATTTATAGTACAAAGCCTACCAGCTGTGAGACAAGAAGTTGAACATGCAATTTT 423
Db 1315 TCNTYTNATHAATAAAGNTAYCARGTCCNCNCNGARYNGYNGCNYTNYTN 1374
Qy 424 GAGCAGCCTCATTTTCGACATCTGTTTGGACATGCTCTGCGTGAAGCGATAATCGGCGT 483
Db 1375 CARCAYCCTTAYATHGARGAYGCGNGTNGCNGGNTCCNGAYGARGTNGCNGNGAY 1434
Qy 484 GCGCGGGCGTCAGTTGATGATGGAAGAAGTCAGTGAGGATATCGTAATTTGGCTGTG 543
Db 1435 YTNCCNGGCGTNGTNGTNGTNYTNAARGAGGNAARTCNATHACNGARAARGARATHCAR 1494
Qy 544 GAATATGTCATAGCTGAGCTGCCGCTTTTATCGGAGCTCCGGATATTTTAGAGGTGAA 603
Db 1495 GAYTAYTNGCNGCARGTACGTCNACNTCTCNAAARAATYTNMGNGCGNGTGNGARTTYGTN 1554
Qy 604 GAGACACTCC 613
Db 1555 AARGARGTNC 1564

RESULT 14
LOCUS AY294513/c 1056 bp DNA linear VRT 28-FEB-2004
DEFINITION Aerodramus terraereginae specimen-voucher DHC20 NADH dehydrogenase
ACCESSION AY294513
VERSION AY294513.1 GI:37812896
KEYWORDS
SOURCE mitochondrion Aerodramus terraereginae
ORGANISM Aerodramus terraereginae
REFERENCE 1 (bases 1 to 1056)
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE The evolution of echolocation in swiftlets
JOURNAL J. Avian Biol. 35 (2), 135-143 (2004)

Db 1495 GAYTAYTNGCNGCARGTACGTCNACNTCTCNAAARAATYTNMGNGCGNGTGNGARTTYGTN 1554
Qy 604 GAGACACTCC 613
Db 1555 AARGARGTNC 1564

RESULT 15
LOCUS AY294501/c 1064 bp DNA linear VRT 28-FEB-2004
DEFINITION Aerodramus vanikorensis specimen-voucher DHC01 NADH dehydrogenase
ACCESSION AY294501
VERSION AY294501.1 GI:37812874
KEYWORDS
SOURCE mitochondrion Aerodramus vanikorensis
ORGANISM Aerodramus vanikorensis
REFERENCE 1 (bases 1 to 1064)
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE The evolution of echolocation in swiftlets
JOURNAL J. Avian Biol. 35 (2), 135-143 (2004)
REFERENCE 2 (bases 1 to 1064)
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2003) Center for Biodiversity, Illinois Natural
History Survey, 607 East Peabody Drive, Champaign, IL 61820, USA
FEATURES
source Location/Qualifiers
1..1064
/organism="Aerodramus vanikorensis"
/mol_type="mitochondrial"
```

```
2 (bases 1 to 1056)
Price,J.J., Johnson,K.P. and Clayton,D.H.
Direct Submission
Submitted (09-MAY-2003) Center for Biodiversity, Illinois Natural
History Survey, 607 East Peabody Drive, Champaign, IL 61820, USA
FEATURES
source Location/Qualifiers
1..1056
/organism="Aerodramus terraereginae"
/mol_type="mitochondrial"
/mol_type="genomic DNA"
/specimen_voucher="DHC20"
/db_xref="taxon:46487"
37..>1056
/gene="ND2"
37..>1056
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAR04261.1"
/db_xref="GI:37812897"
/translation="MNPYAKLIFTMSLLGTMTWISSNHWMAWTGLEINTLAIIBLI
AKPHPRATIEAAIKYVLQAAATLLLFSSINAWSTGQMDITQLANPISCIILLTAI
AMKLGIVPFHPFPEVLQSSLTAMLSLTMKPPPLTILFMTSPSLDTLLTMAIS
STALGGMGLNTOIRKILAFSSISHLGWMAIILYSPKLTLLTFMYCLMTVTVFLT
LNTTKLSTMTWTSWAKAPMLNTTLMMTLLSLAGLPLTGPLPKLIIOELTKQEMT
TAATAIILSLGLFYLRLAYSTITLPPNPNNHMKQWHIYKPTNTLISVLSLSVL
LLPLSP"

Qy 585 GGATATTTTAGAGCTCGAAGACACTCCTTGTCTATCATCTGCGTGGAGCTGGTGA 644
Db 639 GGTAGTTAGGCTCTAGATGAGGATGTTGCTATTCATCTCTAGTGGGAAATGGATGA 580
Qy 645 GAAGATCAGTAACCATGAATTTCTATTTGATGCGCGCGCAATCAAGGTATCTCATTTG 704
Db 579 GAAGGCCAGGATTTTCGGATTTGTTGTTGTTAGTCTATTCAACCCCTAAGGCAGT 520
Qy 705 ACAGGAAATGGC 716
Db 519 AGAAGAGATGC 508

RESULT 15
LOCUS AY294501/c 1064 bp DNA linear VRT 28-FEB-2004
DEFINITION Aerodramus vanikorensis specimen-voucher DHC01 NADH dehydrogenase
ACCESSION AY294501
VERSION AY294501.1 GI:37812874
KEYWORDS
SOURCE mitochondrion Aerodramus vanikorensis
ORGANISM Aerodramus vanikorensis
REFERENCE 1 (bases 1 to 1064)
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE The evolution of echolocation in swiftlets
JOURNAL J. Avian Biol. 35 (2), 135-143 (2004)
REFERENCE 2 (bases 1 to 1064)
AUTHORS Price,J.J., Johnson,K.P. and Clayton,D.H.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2003) Center for Biodiversity, Illinois Natural
History Survey, 607 East Peabody Drive, Champaign, IL 61820, USA
FEATURES
source Location/Qualifiers
1..1064
/organism="Aerodramus vanikorensis"
/mol_type="mitochondrial"
/mol_type="genomic DNA"
```

```
/specimen_voucher="DHC01"
/db xref="taxon:243317"
25...>1064
/gene="ND2"
25...>1064
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAR04251.1"
/db xref="GI:37812875"
/translation="MNPYAKLIPTMSLLCTTWTISSNHVTAAGLEINTLAIIPLI
AKHPHRAIRAAIKYFLVQAASTLLLFSSINAMSTGQWDITQLNPNVSCILLTAL
AMKLGVPFFHFPVEVLQGSPLFTAMLLSTVMKPPPTILFMTSPSLNPTLLATMAIS
STALGWMGLNQTKIRKILAFSSISHLGWMAIILIYSPKLTLLTFMYCMLTITVFLT
LNTTKTLKSTMTSTWTKTLMLSSTLMTLLSLAGLPPLTGFLPKMLIIQELTKQEMT
AAATVISILSLGLFFYLRLAYYSTITLPPNPTNHMKQWHIYKPTNTLISILTSLSIL
LLPLSPMLSSST"

ORIGIN
Query Match      5.4%; Score 40.8; DB 5; Length 1064;
Best Local Similarity 56.8%; Pred. No. 1.7;
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 585 GGATATTTTAGAGTGAAGAGACACTCCTTGCTTATCATCGTCCGTGGAGCTGGGTGA 644
Db 627 GGTCAATTTGGGGCTGTAATAGGATGCTTATTCATCCGAGGTGGGAGATGGATGA 568

Qy 645 GAAGATCAGTAACCATCAATTTCTATTGTATGCGGCGCAATCAAGGTATCTCATTTGT 704
Db 567 GAAGGCTAGGATCTTTCGGATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 704

Qy 705 ACAGGAAATGCG 716
Db 507 AGAAGAAATTGC 496
```

Search completed: December 10, 2004, 16:12:29
Job time : 3355.71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:25:54 ; Search time 18.7485 Seconds
(without alignments)
1288.126 Million cell updates/sec

Title: US-10-627-124-8
Perfect score: 1289
Sequence: 1 MNETGVTERRSVHFAIA.....YLIVQEMAQLSEKRTSEG 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1273	98.8	248	2 A70044	conserved hypotet
2	215	16.7	289	2 G70730	hypothetical prote
3	102.5	8.0	608	2 T02299	hypothetical prote
4	98.5	7.6	332	2 H90174	hypothetical prote
5	94.5	7.3	3418	1 G02334	chromosome segrega
6	93	7.2	1156	2 E69444	breast cancer tumo
7	92	7.1	393	2 A71154	hypothetical prote
8	90	7.0	198	2 T01340	hypothetical prote
9	90	7.0	1198	2 D96723	hypothetical prote
10	89.5	6.9	134	2 S14947	2S albumin - Brazi
11	89.5	6.9	656	2 T52064	dnad-like protein
12	89.5	6.9	898	2 E96659	hypothetical prote
13	89	6.9	420	2 G95005	phosphoribosylamin
14	87.5	6.8	977	2 E86349	hypothetical prote
15	87	6.7	355	2 AE3008	monooxygenase limp
16	87	6.7	335	2 H98275	mtaG protein (Af18
17	87	6.7	365	1 A26522	3-isopropylmalate
18	87	6.7	951	2 G82965	conserved hypotet
19	86.5	6.7	255	2 S66068	conserved hypotet
20	85	6.6	420	2 D97878	phosphoribosylamin
21	85	6.6	716	2 G64816	probable ATP-depen
22	85	6.6	791	2 H72552	hypothetical prote
23	85	6.6	1172	2 C70619	probable lyxX prot
24	84.5	6.6	234	2 F96620	hypothetical prote
25	84	6.5	299	2 G90677	cyn operon positiv
26	84	6.5	299	2 G85528	cyn operon positiv
27	84	6.5	716	2 E90738	probably ATP-depen
28	84	6.5	716	2 G85588	probably ATP-depen
29	84	6.5	818	2 S62790	mismatch DNA recog

probable oligopept
hypothetical prote
probable heavy-met
aspartate transami
probable helicase
probable gamma-glu
conserved hypotet
conserved hypotet
beta-N-acetylhexos
aldehyde ferredoxi
Mx protein - pig
conserved hypotet
hypothetical prote
hypothetical prote
hypothetical prote
cyn operon regulat

ALIGNMENTS

RESULT 1
A:0044
conserved hypothetical protein ymc - Bacillus subtilis
C:0044
conserved hypothetical protein ymc - Bacillus subtilis

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: A70044

R;Kunst, F.; Ogaawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choa
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle
Kieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekiguchi, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: A70044

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-248 <KUN>

A;Cross-references: UNIPROT:O34351; GB:299121; GB:299122; GB:AL009126; NID:G2636029; PIDN

A;Experimental source: strain 168

C;Genetics:

A;Gene: ymc

Query Match 98.8% Score 1273; DB 2; Length 248;
Best local similarity 100.0% Ref. No. 2.7e-101;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	4	MTGVTERRSVHFAIAALTNCEIFERRRHVLVGISPFNSRFSSEDIYRLIGWAKAQFK	63
Db	1	MTGVTERRSVHFAIAALTNCEIFERRRHVLVGISPFNSRFSSEDIYRLIGWAKAQFK	60
Qy	64	SVSVLLAGHEAANLLEALGTPRGAERKVRKVRNRRAERLVAHGGDPKAIHTSFDF	123
Db	61	SVSVLLAGHEAANLLEALGTPRGAERKVRKVRNRRAERLVAHGGDPKAIHTSFDF	120
Qy	124	IDNKAYQLLRQVEHAFPEQPHFRHACLDSRAITGRARGVSLMMEVEEDMLNLAVEY	183
Db	121	IDNKAYQLLRQVEHAFPEQPHFRHACLDSRAITGRARGVSLMMEVEEDMLNLAVEY	180
Qy	184	VIAELEFFICAPDILEVEETLLAYHRPWKLGEKISNHFSCIMRPNQGYLIVQEMAQMLS	243
Db	181	VIAELEFFICAPDILEVEETLLAYHRPWKLGEKISNHFSCIMRPNQGYLIVQEMAQMLS	240
Qy	244	EKRITSEG	251

Db 241 EKRIITSEG 248
|||||||
RESULT 2
G07030
hypothetical protein RV2275 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: G70730
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70730
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-289 <COL>
A;Cross-references: UNIPROT:Q50688; GB:Z77163; GB:AL123456; NID:g3261610; PIDN:CAB00960.
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: RV2275
Query Match 16.7%; Score 215; DB 2; Length 289;
Best Local Similarity 27.3%; Pred. No. 7.2e-11;
Matches 62; Conservative 34; Mismatches 115; Indels 16; Gaps 4;
QY 16 FTAEALTENCREFERRRRVHLVGISPFNSRSEDYIYRLIGWAKAFKSVSVLLAGHEAA 75
Db 64 FLVVRPTQCCQIHTEGDHAVIGVSCNSYFSRQLRDLGLGNTFDRVDFVYTDVHVA 123
QY 76 NLLEALGTPRGKAEKVRKEVSRNRFAERALVAHGDGPK-----AHTFSDFDIDNKAYQL 131
Db 124 ESYEALGDSAIEARRKAVKNIRGV--AKITTTVNEILDPA GARLCVRPNSFQSQNEAYRE 181
QY 132 LROVEVHAPEQPHFRHACLDMSREAL--IGRARGVSLMMEVSEDMNLNLAVEYVIAEL 198
Db 182 LHADLLTRLKDDBLRAVCDLVRFLSTKVGPRQGATATQEQV-----CMDYICAEA 234
QY 189 PFTIGAPDILEVETLLAYHRPWLKGEKISNHEFSICMRPNOCYLIV 235
Db 235 PLFLDPTAILGVPSLLNCHQSLPLAEMLYARGSLGRASRNOQHAIV 281
RESULT 3
T02299
hypothetical protein F23858.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C;Accession: T02299; T46353
R;Lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; G.
Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Attix, C.; Andreise, T.; Trankhei
submitted to the EMBL Data Library, March 1998
A;Authors: Duarte, S.; Lucas, S.; Bruce, R.; Thomas, P.; Quan, G.; Kronmiller, B.; Arell
A;Description: Sequence analysis of a human P1 clone containing the XRCC9 DNA repair gen
A;Reference number: Z14637
A;Accession: T02299
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-608 <LAM>
A;Cross-references: UNIPROT:O60378; EMBL:AC004475; NID:g29988396; PIDN:AAC08052.1; PID:g2
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23037
A;Accession: T46353
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 217-608 <AAA>
A;Cross-references: EMBL:AL137286
A;Experimental source: adult testis; clone DKFp434E2216

C;Genetics:
A;Map position: 19
A;Introns: 58/1; 84/1; 143/3; 185/1; 218/3; 260/1; 378/3; 414/2; 473/3; 509/2; 558/1; 601
A;Note: DKFp434E2216.1
Query Match 8.0%; Score 102.5; DB 2; Length 608;
Best Local Similarity 26.3%; Pred. No. 0.73;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;
QY 84 PRGKAERKVRKEVSRNRFAERALVAHGDGPKAIHTFSDIDNKAYQLLRQEVHAFPEQ 143
Db 146 PEGAEATRKVIEKLAR-----FVAEGGPELEKVMEDYKPNPAFLHDKNSRFL-- 195
QY 144 PFRHACLDMSREALIIGRARGVSLMMEVSEDMNLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 196 -YRKVAIRKEAQKSAQSVSPPE--DEEVKNLA-----EKLARFIADGGP---EVE 245
QY 202 ETLLAYHRPWLKGEKISNHEFSICMRPN--QGYLIVQEMAQMLSEKRITSEG 251
Db 246 TIALQNNR-----ENQAFGLYEPNSQGYKYRQKLEEFPRKAKASSTG 288
RESULT 4
H90174
hypothetical protein SSO0320 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: H90174
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-v
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: H90174
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-392 <KUR>
A;Cross-references: UNIPROT:Q980H7; GB:AE006641; NID:g13813463; PIDN:AAK40655.1; GSPDB:G5
C;Genetics:
A;Gene: SSO0320
Query Match 7.6%; Score 98.5; DB 2; Length 392;
Best Local Similarity 26.1%; Pred. No. 0.93;
Matches 65; Conservative 35; Mismatches 84; Indels 65; Gaps 14;
QY 8 VTERRSVHFIABALTENCREF-ERRRHVLVGI-----SPNSRPSDEY--IYRLIGW 57
Db 83 VIESKOLKFTIENKL----RELLDEYKRRIYKGIKLEGYTPDPPIETPTSIYETINKVIGG 138
QY 58 AKAQFKSVSVLL-----AGHEAANLLEALGTPRGKAERKVRKEVSRN 99
Db 139 YEADTSMIEVYLYTWNVYKEYLAKDKIKCSDDDVARYLLE-----GKGKIKIDKSSAN 192
QY 100 RRAERALVAHGDGPKAIHTFSDIDNKAYQLLRQEVHAFPEQPHFRHACLDMSREALI 159
Db 193 -----VSMLVNYG--VTSVQKLLDFIDINGSEVYRSKKET-----KGILLD---EALA 236
QY 160 GRARGVSLMMEVSEDMNLNLA---EYVIAELPFFIAGPDILEVE--TLAYHRPWLKGE 215
Db 237 GNAL-VSEYERTINEEYLSIAIKVADYIKNNLQHERGFRDIKIDNITKVPYLEP----- 290
QY 216 KISNHEFSI 224
Db 291 -ISNSEASI 298
RESULT 5
G02334
breast cancer tumor suppressor BRCA2 - human
N;Alternate names: breast cancer susceptibility protein BRCA2
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004

C;Accession: G02334; S68501
R;Tavtigian, S.V.; Rommens, J.M.; Couch, F.J.; Neuhausen, S.; Bell, R.; Berry, S.; Bogden, M.; Snyder, S.; Stringfellow, M.; Stroup, C.; Swedlund, B.; Teng, D.; Thomas, A.; et al. submitted to the EMBL Data Library, December 1995
A;Reference number: H01078
A;Accession: G02334
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3418 <TAV>
A;Cross-references: UNIPROT:P51587; EMBL:U43746; NID:g1161383; PIDN:AAB07223.1; PID:g1161383
R;Wooster, R.; Bignell, G.; Lancaster, J.; Swift, S.; Seal, S.; Mangion, J.; Collins, N.; Ith, A.; Connor, F.; Arason, A.; Gudmundsson, J.; Ficenec, D.; Kelbell, D.; Ford, D.; et al.; Harad, S.; Lenoir, G.; Egilsson, V.; Barkadottir, R.B.; Easton, D.F.; Bentley, D.R. Nature 378, 789-792, 1995
A;Authors: Futreal, P.A.; Ashworth, A.; Stratton, M.R.
A;Title: Identification of the breast cancer susceptibility gene BRCA2.
A;Reference number: S68501; MUID:96112016; PMID:8524414
A;Accession: S68501
A;Molecule type: mRNA
A;Residues: 282-371, 'N', 373-598, 'S', 600-1108, 'EQ', 1111-1119, 'D', 1121-2321, 'V', 2323-2386, 'C';
A;Gene: GDB:BRCA2
A;Cross-references: GDB:387848; OMIM:600185
A;Map position: 13q12.3-13q12.3
C;Superfamily: DNA recombination repair protein, BRCA2 type
C;Keywords: polymorphism; tumor suppressor

Query Match 7.3%; Score 94.5; DB 1; Length 3418;
Best Local Similarity 18.4%; Pred. No. 31;
Matches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;
Mismatches 53; Conservative 57; Mismatches 109; Indels 69; Gaps 11;

Query 8 VTRRSVHFAEALTENCRIFFERRHV-----LVGISPFNSRSEDIYVRLIG 56
Db 2109 VDRNPEHCNVSEMEKCTCKEFKLSNLLNVEGGSSNNHSIKVPSYLSFQDQKQQLVLG 2168

Query 57 WAKAQFQSVSVLLAGHAANLLEALGTPRGAERKVRKEVSRNRRFAERALVAHGGDPKA 116
Db 2169 TKVSLVENIHVL--GKEQA-----SPKNVMEIGKTETTFSD---VPVKTNIIEV 2211

Query 117 IHTFS-----FDINKAYQLLRQVE-----HAFEPQPHFRHACLDMSREAI 158
Db 2212 CSTYSKDSSENYFTEAVEIAKAPWDELDTSKLPSHATSLTFCPENEEVLSNSR--- 2268

Query 159 IGRARGVSLMM-----EEVSEMDLN---LAVRYIAELPFFICAPDILEVEETLLAYH--- 208
Db 2269 IGRGEPPLIVGEPSIKRNLNLEFDRIENQEKSLKASKSTPDGTIKDRRLFMHVSLE 2328

Query 209 -----RPKLGKESINHFSCMRPNQGLIVQEAQMLSEKRITS 249
Db 2329 PITCVPRFTTKRQEQINPNFT---APQBFLSKSHLYHTLEKSSS 2373

RESULT 6
E69444
Chromosome segregation protein (smc1) homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: E69444
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Arttich, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: E69444
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1156 <KLE>
A;Cross-references: UNIPROT:O28714; GB:A6000995; GB:A6000782; NID:g2689318; PIDN:AAB8969
C;Superfamily: chromosome segregation protein SMC1

Query Match 7.2%; Score 93; DB 2; Length 1156;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 39; Conservative 47; Mismatches 59; Indels 58; Gaps 9;
Mismatches 39; Conservative 47; Mismatches 59; Indels 58; Gaps 9;

Query 1 MNEMTGMVTTERRS-----VHFIAEALTENCRIFFERRRHVLGVISPFNSR 45
Db 707 VDRLTGMISLRNRISILDEKIRTESGRIELEKISQKSR-----KENVISSLDYNSK 762

Query 46 FSEDIYRILIGWAKAQFQSVSVLLAGHAANLLEALGTPRGAERKVRKEVSRNRRFAER 105
Db 763 LAE--MEEAIGEAEETIEERMLRGSEVPKIVBELD-----KIKEEHQRNREI--- 809

Query 106 ALVAHGGDPKAIHTFSDFDINK--AVQLLRQVEVHAFEPQPHFRHACLD-----MSREAIL 159
Db 810 -LIS-----IEKTIESTEFKREQLSSMQEKQVYLDKIDRIDEIRRTIEE 854

Query 160 GRARGVSLMMEEVSEDMNLAVE 182
Db 855 GKAR-----VEEINSELEELRKE 872

RESULT 7
A71154
Hypothetical protein PH0432 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: A71154
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: A71154
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-393 <KAW>
A;Cross-references: UNIPROT:O58169; GB:AP000002; NID:g3236129; PIDN:BAA29518.1; PID:d1031
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0432

Query Match 7.1%; Score 92; DB 2; Length 393;
Best Local Similarity 21.2%; Pred. No. 3.4;
Matches 56; Conservative 40; Mismatches 92; Indels 76; Gaps 11;

Query 30 ERRRHVLGVISPFNSRF--SEDIYRILIGWAKAQFQSVSVLLAGHAANLLEALGTPRGK 87
Db 5 EKPRIAIIIGURGPSKYGGTETTFVELTSRLKDAFK-----FVNMHETKTFE----- 52

Query 88 AERKVRKEVSRNRRFAERALVAHGGDPKAIHTFSDFDINKAYQLLRQVE--VEHAFEPQPH 145
Db 53 -----DEYNRIIRVHSPAIESKSTSIPIINDFT----NTAYMLANHEKIDIELFYFLGP- 101

Query 146 FRHACLDMSREAIIGRARGVSLMMEEVSEDMNLMLA-----GEKISNHFESI-----VE 182
Db 102 -----DSSLAAILARLGRKKVLINPDGVWRRLIKRSYFVPFYLPIFYFATIIMYFME 155

Query 183 YVIAELPFFICAPDILEVEETLLAYHHPWKL-----GEKISNHFESI-----CM 226
Db 156 YLSCKLPDIVVA--DSGIKHEHLEKRRKPRRVVITYGARELISSEFSVESEKEILSRFNL 214

Query 227 RPNQGLIVQEAQMLSEKRITSE 250
Db 215 EPLGYLTV---ARIVAENNIHME 235

RESULT 8
T01340
Hypothetical protein F6N15.18 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C;Accession: T01340

R; Ryan, E.; Edwards, J.; Pape, K.
submitted to the EMBL Data Library, May 1998
A; Description: The sequence of A. thaliana F6N15.
A; Reference number: Z14297
A; Accession: T01340
A; Status: translated from GB/EMBL/DBJ
A; Molecule type: DNA
A; Residues: 1-198 <RNA>
A; Cross-references: UNIPROT:O81313; EMBL:AF069299; NID:g3193311; PID:g3193315
A; Experimental source: cultivar Columbia
C; Genetics:
A; Map position: 4
A; Note: F6N15.18

Query Match 7.0%; Score 90; DB 2; Length 198;
Best Local Similarity 29.3%; Pred. No. 2.1;
Matches 39; Conservative 22; Mismatches 46; Indels 26; Gaps 8;

QY 97 SRNRFFAEALVAHGDPKAIHTFSDFDINKAVQLLRQEVHAFFEQPHRFHACLMSRE 156
||| : : : ||| : : : ||| : : : ||| : : : |||
Db 20 SRSNHSPKRSMW----EPQPHLLMDW--NKANDLLTQE-HAAFLNDPH--HLMLDPPPE 70
||| : : : ||| : : : ||| : : : ||| : : : |||

QY 157 AITGRARGVSIMMEVESEDLNL-AVEYVTAEL-----PFFIGAPD-----ILEVEETL 204
||| : : : ||| : : : ||| : : : ||| : : : |||

Db 71 TLIHLDDE-----EYEDDMAMKEQMIAVQPVDIDFATVPKPNRRNVRISSDDPTV 125
||| : : : ||| : : : ||| : : : ||| : : : |||

QY 205 LAYHRPWKLGEKI 217
: | : : : |||
Db 126 VARRRERISEKI 138
: | : : : |||

RESULT 9
D96723
hypothetical protein F20P5.20 [imported] - Arabidopsis thaliana
C; Species: Arabidopsis thaliana (mouse-ear cress)
C; Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C; Accession: D96723
R; Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A; Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C. Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzalli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A; Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A; Reference number: A86141; MUID:21016719; PMID:11130712
A; Accession: D96723
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-1198 <STO>
A; Cross-references: UNIPROT:C04538; GB:A8005173; NID:g2194131; PID:N.AAB61106.1; GSPDB:GNC
C; Genetics:
A; Map position: 1

```

QY 192 IGAPDILE-VEETLAY 207
      : : : : :
Db 760 LALKPLKMGEEGNLPF 776

RESULT 10
SI4947
2S albumin - Brazil nut
C:Species: Bertholletia excelsa (Brazil nut)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: SI4947
R:Gander, E.S.; Holmstrom, K.O.; de Paiva, G.R.; de Castro, L.A.B.; Carneiro, M.; Gross, J.
Plant Mol. Biol. 16, 437-448, 1991
A:Title: Isolation, characterization and expression of a gene coding for a 2S albumin from
A:Reference number: SI4946; MUID:91370890; PMID:1840683
A:Accession: SI4947
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <GAN>
A:Cross-references: EMBL:X54491; NID:gl7712; PIDN:CAA38363.1; PID:gl7713
C:Genetics:
A:Introns: 67/3
C:Superfamily: wheat alpha-amylase inhibitor

      Query Match          6.9%; Score 89.5; DB 2; Length 154;
      Best Local Similarity 20.5%; Pred. No.1.7;
      Matches 27; Conservative 25; Mismatches 41; Indels 39; Gaps 3;

QY 54 LIGWAKQFKSVSVLLAGHAEANLLLEALGTPRGKAERKVRKSVNRNRFARALVAHGSD 113
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 VLGQATAFRITVTITLLEEQEN-----PGRSEQQCRQMERQQQL----- 57

QY 114 PKAHTFTSDFDINKAYQILLRVEVEHAF-----EOPHFRHACLDMSREAIIGRARG 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 -----NHCRMVLRQOMESPYQNPLRGRGEPHLDCECEQLERMDMCRCEG 105

QY 165 VSLAMVEEVSDEM 176
      : : : : :
Db 106 LRMLLRQREEM 117

RESULT 11
TS2064
dnaJ-like protein [imported] - rice
C:Species: Oryza sativa (rice)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004
C:Accession: TS2064
R:Illaca, V.; Lou, A.; Young, S.; Messing, J.
submitted to the EMBL Data Library, December 1998
A:Description: Analysis of a gene-dense region in Oryza sativa.
A:Reference number: 225930
A:Accession: TS2064
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-656 <LIA>
A:Cross-references: UNIPROT:Q9XEM8; EMBL:AF111710; PIDN:AAD27555.1
A:Experimental source: subsp. indica
C:Genetics:
A:Map position: 5
A:Introns: 46/2; 125/3; 190/1; 235/3; 268/1; 308/3; 328/3; 356/2; 421/3; 465/1; 489/1; 54/

```

Qy	19	EALTE	KREIFERRRHVLVIGSPNRSFSEY--IYRLIGWAKAOFKSVSLLAGHEAA	76
Dd	616	EGAECC	KLVF-----AGVKPLVSQTASGVMLNLVAGSKVTRKS-----SCTEAGK	663
Qy	77	LLEALGT	PRGKAERKVRKE---VSRNRFRFAELVALVAHGGDPKAIHTFSDFIDNKAYOLL	132
Dd	664	VLA-Q-	GRSLEEAKKLVEKSGFYGYSSNVTVAAKQLAE-----IDNK-IEIL	708
Qy	133	RQEVEHA	FPEQPHRFACLDMSREALIGR-ARGVSLMMEVEESDMNLNAVEVVIATLPPF	191
Dd	709	SSETSD	EALDK-----KSKLLSARDKYETIVLKEELREEKKAEQRRRMELBRF	759

```
Query Match      6.9%; Score 89.5; DB 2; Length 656;  
Best Local Similarity 23.7%; Pred. No. 10;  
Matches         44; Conservative 32; Mismatches 61; Indels 49; Gaps 10;
```

QY		42	FNSRPFSEYYIRLIOWAKAQFQSVSV--LLAGHEAANILLEALCTPRGKAERKVRKEVERSN	99
	:	:	: : :	: : :
Db		381	FGSDYFDYVGQL---ALASIASVEEENLNGQA-----RGKVQEKI-KELQKE	426
	:	:	:	:
QY		100	RRFARERALVANGGDPKATHTPSD-----FIDNKAYQLLR-----QEVEHAFFPQP----	144
	:	:	:	:
Db		427	R-----EQGL:QSLKDRIQQPFFVDREDFNVNWANGSAORLSHAFAGEAMLTTI	474
	:	:	:	:

Qy 145 ---HFRHACLDGREAL--IGRARGVSLMWE-EVSDMLNLAVEYVIAELPPFIGAPDIL 198
Db 475 GYIVYVQAARELKSALALQLOEGMKMBESDNKEDQIMKSPEEKDAMLVYLKINNV 534
Qy 199 EVEETL 204
Db 535 DIESTL 540

RESULT 12
E96659
hypothetical protein F9N12.3 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: E96659
R;Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E96659
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-898 <STO>
A;Cross-references: UNIPROT:Q9C8T9; GB:AE005173; NID:G997185; PIDN:AAF34847.1; GSPDB:GN
C;Genetics:
A;Map position: 1
C;Superfamily: disease resistance protein RPS2; leucine-rich alpha-2-glycoprotein repeat

Query Match 6.9%; Score 89.5; DB 2; Length 898;
Best Local Similarity 20.1%; Pred. No. 15;
Matches 57; Conservative 49; Mismatches 84; Indels 93; Gaps 13;
Qy 1 MNEMTGWYTER---RSVHFIAEALTENCREFERRERHVLGVSPNSRFS--EDYIYR 53
Db 333 IRELSRVVAKCCGLPALNVSE--TMSCKRTVQEWRAHYVLNVAAPSGMDDKILP 390
Qy 54 LIGWAKAQFQSVSVLLAGHEAANLLEALGTPRGKAERKVRKVRNRRRFAERALVAHGGD 113
Db 391 LLKXSYDSLKGDVQWC-----LLYCALPFE---DAKIRKENLIEWICEIIDGSEGI 441
Qy 114 PKAIHTSDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSRAIIGRARGVSLMMEVS 173
Db 442 DKA-----ENQGYE-----IIGSLVRASILLMEVEE 466
Qy 174 EDMLNL-AVEYVIAELPPFI-----GAPDILEVEETLLAYHRPWKLGE 215
Db 467 LGGANIVCLHDVREMAWLTASDLGKQNEAFIVRASVGLREILKVEN-----MNVVR 518
Qy 216 KISNHEPSICWRPN-----OGVLIYQEAQMLSE-----KRITSE 250
Db 519 RMS-----LMKNNIAHLDRLCDMELTLLQSTHLEKISSE 555

RESULT 13
G95005
phosphoribosylamine-glycine ligase [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: G95005
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: G95005
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-420 <KUR>
A;Cross-references: UNIPROT:Q97T98; GB:AE005672; PIDN:AAK74240.1; PID:gl4971515; GSPDB:GN
C;Genetics:
A;Gene: SP0051
C;Superfamily: phosphoribosylamine-glycine ligase; phosphoribosylamine-glycine ligase hom
Query Match 6.9%; Score 89; DB 2; Length 420;
Best Local Similarity 23.2%; Pred. No. 6.6;
Matches 57; Conservative 34; Mismatches 89; Indels 66; Gaps 10;
Qy 15 HFTAEALTE--NCREIFFERRR---VLVGVSPNSRFSYDIYRLIGWAKAQFQSVSL 69
Db 13 HATAKKLESKQVEKVFVAPGNDGMTLDGLVNLISSEH--YKLIDFAKT--NDVAWTF 68
Qy 70 AGHE---ANL-----LEALGTPRGKAERKVRKVRNRRRFAERALVAHGGDKPAIH 118
Db 69 IGPDALLAGIVDDFNQAGLKAFGPTRAAAALEWSKD-----FAKEIMVKYGVPTATYG 122
Qy 119 TFSDFIDNKAYQLLRQVEHAFPEQPHFRHACLDMSREALIGRARGVSL---MMEEVSE 174
Db 123 TFSDFBEAKAY-----TEKHGAPIVVRADGALGKGVVVAETVE 161
Qy 175 DMLNLAVEYVIAELPPFI GAPDILEVEETLLAYHRPWKLGEKISNHEFSICMRPNQGYLI 234
Db 162 QAVEAAHEMLLDNKGDSGARVIE-----EFLEGEFEFSLFAFVNGDKFY 206
Qy 235 VQEMAQ 240
Db 207 IMPTAQ 212

RESULT 14
E86349
hypothetical protein F8K7.6 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: E86349
R;Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E86349
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-977 <STO>
A;Cross-references: UNIPROT:Q9XI14; GB:AE005172; NID:G5263315; PIDN:AAD41417.1; GSPDB:GN
C;Genetics:
A;Map position: 1
C;Superfamily: preprotein translocase secA
Query Match 6.8%; Score 87.5; DB 2; Length 977;
Best Local Similarity 22.8%; Pred. No. 25;
Matches 62; Conservative 50; Mismatches 85; Indels 75; Gaps 19;
Qy 1 MNEMTGWYTER---SVHFIAEALTENCREFERRRHLVGLSPNSRFSYDIY--- 52
Db 385 INELTGRVEDKRWRSEGVHQAVEA---KEGLEIQADSIWVAQITQSLFK---LYPKL 436
Qy 53 -RLIGWAKAQFK-----SVSVLLAGHEAANL-----LEALGTPRGKAERKVRKVRNRR 100
Db 437 SGMGTGATKEBFLKMFQIPVIEVPTNLSNIRIDLFIQAFATARGKWEH--VRREV--ED 493

```
QY 101 RFAERALVAHGDPKAIHTSFIDNKAY--QLLRQEVHAFPEQPHF-----RHACLD 152
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 494 MFQ-----GRPVLVGTTS--VENSEYLSLLKE-----WGIPHNVLNARPKYAARE 538
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 153 MSREAIIGRARGVSLMMEEVSEMDLNLAVEYVIAELPFFIGAPDILE---VEETLLAYHR 209
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 539 ADFIAQAQGRKYAIT-----ISTNAGRGTDIILG-----GNPKMLAREIIEDSILSYLT 587
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 PWKLGKISNHFPSICMRPNQGYLIIVQEMAQM 241
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 588 SEVLADNIDDELS---QKN-----LINEQSEM 612
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AE3008
monooxygenase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AE3008
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
; Karp, P.; Gillet, W.; Grant, C.; Guenthermer, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE3008
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <KUR>
A:Cross-references: UNIPROT:Q8U9Q5; GB:AE008689; PIDN:AAL44483.1; PID:g17742088; GSPDB:G
C:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3671
A:Map position: linear chromosome
C:Superfamily: alkanal monooxygenase [FMN-linked] (bacterial luciferase)

Query Match 6.7%; Score 87; DB 2; Length 355;
Best Local Similarity 23.5%; Pred. No. 7.9;
Matches 36; Conservative 28; Mismatches 63; Indels 26; Gaps 7;

QY 53 RLIGWAKAQFYSVLLAGHEAANLLEALGTPRKAERKYRKEVSRNRERFAERALVAHGG 112
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 RIFFWACQ-STETFIAGRRGLNVLTA---LGTTESLAPKIAAYRKRERK-----HGH 217
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 DPKA-----IHTFSDPIDNKAYQLLRQEVHAFPEQPHFRHACLDMSREAIIGRARGVS 166
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 DPAAGTVSMNVHIFLGDDTGK-----VKSNSVKPFGDYLRTHYHLEGG-----LGRSMGVD 268
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 LMMEEVSEMDLNLAVEYVIAEL---PPFIGAPD 196
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 ITLDNFSDDDLSLIERGIEGFMKGRSLIGTPE 301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: December 10, 2004, 13:45:52
Job time : 21.7485 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:47:24 ; Search time 96.4207 Seconds
(without alignments)
1497.801 Million cell updates/sec

Title: US-10-627-124-8

Perfect score: 1289

Sequence: 1 MNETGNTTERRSVHFIAE.....YLIVQMAQVLSEKRTISG 251

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1273	98.8	248	2	O34351
2	400.5	31.1	234	2	Q7N9M5
3	249	19.3	239	2	Q8GED7
4	215	16.7	289	2	Q7TVV7
5	215	16.7	289	2	Q50688
6	102.5	8.0	608	2	O60378
7	102.5	8.0	641	2	Q8WMT4
8	102.5	8.0	643	2	Q6P3X9
9	102.5	8.0	643	2	AH63784
10	102.5	8.0	645	2	Q8IWX8
11	100.5	7.8	700	2	Q8XUM4
12	100	7.8	1093	2	Q82527
13	98.5	7.6	392	2	Q980H7
14	97	7.5	864	2	Q75130
15	94.5	7.3	3418	1	BRC2_HUMAN
16	93	7.2	1156	2	O28714
17	92	7.1	393	2	O58169
18	91	7.1	652	2	Q9F531
19	91	7.1	751	2	Q86ZL0
20	90.5	7.0	360	2	Q95XC8
21	90.5	7.0	366	2	O6L1I8
22	90	7.0	916	2	Q94E26
23	90	7.0	1198	2	O04538
24	89.5	6.9	392	2	Q7N1G4
25	89.5	6.9	556	2	Q9SH21
26	89.5	6.9	656	2	Q9XW68
27	89.5	6.9	898	1	DR19_ARATH
28	89	6.9	420	1	PUR2_STRPN
29	88.5	6.9	399	2	Q8CV34
30	88	6.8	285	2	Q9F5V9
31	88	6.8	643	2	Q8CH02

32	88	6.8	646	2	Q8R094
33	88	6.8	646	2	AH27188
34	88	6.8	1058	2	Q74260
35	88	6.8	1058	2	AAS54836
36	87.5	6.8	292	2	Q8A7C1
37	87.5	6.8	895	2	Q89JF5
38	87.5	6.8	975	2	Q8ZWR5
39	87.5	6.8	977	2	Q9X1I4
40	87	6.7	355	2	Q7C1I6
41	87	6.7	355	2	Q8U9Q5
42	87	6.7	365	1	LEU3_BACSU
43	87	6.7	501	2	Q7ZXV7
44	87	6.7	647	2	Q7Y145
45	87	6.7	691	2	Q6WL05

ALIGNMENTS

RESULT 1

ID	O34351	PRELIMINARY;	PRT;	248 AA.
AC	O34351; Q795E6;			
DT	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	01-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	YvmC.			
GN	Names=yvmC; OrderedLocustNames=BSU35070;			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98230327; PubMed=9570401;			
RA	Reizer J., Holscher C., Titgemeyer F., Rivolta C., Rabus R.,			
RA	Stulke J., Karamata D., Saier M.H. Jr., Hillen M.;			
RT	"A novel protein kinase that controls carbon catabolite repression in			
RT	bacteria.";			
RL	Mol. Microbiol. 27:1157-1169 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Lazarevic V., Soldo B., Rivolta C., Reynolds S., Mauel C.,			
RL	Karamata D.;			
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.-Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haitech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M.,			
RA	Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,			
RA	Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,			
RA	Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,			
RA	Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C.,			
RA	Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,			
RA	Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,			
RA	Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,			
RA	Prescott A.M., Prescan E., Pujic P., Purnelle B., Rapoport G.,			
RA	Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,			
RA	Rose M., Sadaie Y., Sato T., Scanlan A., Schleich S., Schroeter R.,			
RA	Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P.,			
RA	Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,			
RA	Takenaru K., Takeuchi M., Tanakoshi M., Tanaka T., Terptrat P.,			
RA	Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,			

Query Match 8.0%; Score 102.5; DB 2; Length 608;

```
Best Local Similarity 26.3%; Pred. No. 8.4;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGAERKVRKVEVSRNRRFAERLVAHGDPKAIHTFSDPIDNKAYQLLRQVEHAFPEQ 143
Db 146 PEGAETRKVIEKLAR-----FVAEGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 195

QY 144 PFRHACLDMSRAIIGRARGVSLMMEVEVSDMLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 196 -YRKVKVAIRKGAQKSAQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 245

QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRIITSG 251
Db 246 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 288

RESULT 7
Q8WWT4 ID Q8WWT4 PRELIMINARY; PRT; 641 AA.
AC Q8WWT4
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2004 (TrEMBLrel. 20, Last sequence update)
DE RNA-binding protein splice variant a.
GN Name=RBP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu Y., Nguyen C.-T.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY072917; AAL68961.1; -.
DR Genew; HGNC:18643; SF4.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR000467; G_patch.
DR InterPro; IPR000061; Surp.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF01805; Surp; 2.
DR PROSITE; PS0174; G_PATCH; 1.
DR PROSITE; PS0128; SURP; 2.
SQ SEQUENCE 641 AA; 72011 MW; 60BCA0B5EB1A0D80 CRC64;

Query Match 8.0%; Score 102.5; DB 2; Length 641;
Best Local Similarity 26.3%; Pred. No. 8.9;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGAERKVRKVEVSRNRRFAERLVAHGDPKAIHTFSDPIDNKAYQLLRQVEHAFPEQ 143
Db 179 PEGAETRKVIEKLAR-----FVAEGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 228

QY 144 PFRHACLDMSRAIIGRARGVSLMMEVEVSDMLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 229 -YRKVKVAIRKGAQKSAQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 278

QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRIITSG 251
Db 279 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 321

RESULT 8
Q6P3X9 ID Q6P3X9 PRELIMINARY; PRT; 643 AA.
AC Q6P3X9;
DT 05-JUN-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE SF4 protein (Fragment).
GN Name=SF4;
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Hopkins R.F., Zebdan K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko L., Marusina K., Falmer A.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC063784; AAH63784.1; -.
DR InterPro; IPR000467; G_patch.
DR InterPro; IPR000061; Surp.
DR Pfam; PF01585; G_patch; 1.
DR Pfam; PF01805; Surp; 2.
DR SMART; SM00443; G_patch; 1.
DR SMART; SM00648; SWAP; 2.
DR PROSITE; PS0174; G_PATCH; 1.
DR PROSITE; PS0128; SURP; 2.
FT NON TER 1
SQ SEQUENCE 643 AA; 72252 MW; 4DE9A061B9C1C2D1 CRC64;

Query Match 8.0%; Score 102.5; DB 2; Length 643;
Best Local Similarity 26.3%; Pred. No. 8.9;
Matches 45; Conservative 24; Mismatches 71; Indels 31; Gaps 8;

QY 84 PRGAERKVRKVEVSRNRRFAERLVAHGDPKAIHTFSDPIDNKAYQLLRQVEHAFPEQ 143
Db 181 PEGAETRKVIEKLAR-----FVAEGGPELEKVMEDYKDNPAFAFLDKNSREFL-- 230

QY 144 PFRHACLDMSRAIIGRARGVSLMMEVEVSDMLNLAVEYVIAELPFFI--GAPDILEVE 201
Db 231 -YRKVKVAIRKGAQKSAQKVSPPPE-DEEVKNLA-----EKLARFIADGGP---EVE 280

QY 202 ETLLAYHRPWKLGEKISNHEFSICMRPN-QGYLIVQEMAQMLSEKRIITSG 251
Db 281 TIALQNNR-----ENQAFSLYEPNSQGYKYRQKLEEFKAKASSTG 323

RESULT 9
AAH63784 ID AAH63784 PRELIMINARY; PRT; 643 AA.
AC AAH63784;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DE SF4 protein (Fragment).
GN SF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```



```
Db 233 HEAARVLOTLSFAPQHQAALATARANVLKRVRLDEALAVARQAVVLAPRSAAEHALLA 292
Qy 122 DFDNKAQYLLRQVEHAFPEQHERHACL---DMSREALIGARGVSLMMVEEVSMDLN 178
Db 293 -----MALQTLGTDE-----ALPHFQOARLPQAVAEALVGR-----TLMEAGRDAAR 340
Qy 179 LAVYVIAELP-----FFIGAPDILEVEETLLAYHRPWKLGKISNHEFSI 224
Db 341 AAFDQALEQPGSVQALAGRADARTTAGDPDIALEACLAEGERR-SLRDRISAH-FAL 398
Qy 225 CMRPNOGYLIVQEMAQ 240
Db 399 -----GRAYLDLQDPAR 410

RESULT 12
ID Q82527 PRELIMINARY; PRT; 1093 AA.
AC Q82527;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Putative lysyl-tRNA synthetase.
GN Name=lys2; OrderedLocusNames=SAV7296;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
CC -I- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
DR EMBL; AP005050; BAC75007.1; -.
DR HSSP; P14825; IE10.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004815; F:aspartate-tRNA ligase activity; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004824; F:lysine-tRNA ligase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006422; F:aspartyl-tRNA aminoacylation; IEA.
DR GO; GO:0006430; F:lysyl-tRNA aminoacylation; IEA.
DR InterPro; IPRO07424; DUF470.
DR InterPro; IPRO07425; DUF471.
DR InterPro; IPRO07426; DUF472.
DR InterPro; IPRO08994; Nucleic_acid_OR.
DR InterPro; IPRO04364; tRNA-synt_2.
DR InterPro; IPRO02312; tRNA-synt_2.
DR InterPro; IPRO02313; tRNA-synt_lys_2.
DR InterPro; IPRO04365; tRNA anti.
DR InterPro; IPRO06195; tRNA_ligase_II.
DR Pfam; PF04329; DUF470; 1.
DR Pfam; PF04330; DUF471; 1.
DR Pfam; PF04331; DUF472; 1.
DR Pfam; PF00152; tRNA-synt_2; 1.
DR Pfam; PF01336; tRNA anti; 1.
DR PRINTS; PRO1042; TRNASYNTHASP.

DR PRINTS; PRO0982; TRNASYNTHLYS.
DR TIGRFAMS; TIGR00499; lys_bact; 1.
DR PROSITE; PS50862; AA_TRNA_LIGASE_II; 1.
KW ATP-binding; aminoacyl-tRNA synthetase; Complete proteome; Ligase;
KW Protein biosynthesis.
SQ SEQUENCE 1093 AA; 119034 MW; B916D8B13D6ABE53 CRC64;

Query Match 7.8%; Score 100; DB 2; Length 1093;
Best Local Similarity 26.5%; Pred. No. 26;
Matches 36; Conservative 15; Mismatches 47; Indels 38; Gaps 4;

Qy 19 EALTENC-----REIFERRRHVLVGISPFNSRFSFSEDIYIRLIGWAKAQ 61
Db 939 EALHRRHCDRAGVPYTTADDCRGDWLEMYERLVEERTGLTFYKDFPTD----- 986
Qy 62 FKSVSVLLAGHEA-ANLLE-----ALGTPRGKAERKVRKESVNRFPFAERLVAHGGDPK 115
Db 987 ---VSLPTQHRADPRLAERWDLVAFGTGLTAYSELTPMEQRRLTAQSLLAAGDPE 1043
Qy 116 AJHTSFDSFDIDNKAYQL 131
Db 1044 AMELDEDFDUALEYAM 1059

RESULT 13
ID Q980H7 PRELIMINARY; PRT; 392 AA.
AC Q980H7;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=SSO0320;
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Chauvo G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AS006666; NAK40655.1; -.
DR FIR; H90174; H90174.
DR InterPro; IPRO08928; Glyco_trans_6bp.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 392 AA; 44603 MW; F7E9CBDDA820C556 CRC64;

Query Match 7.6%; Score 98.5; DB 2; Length 392;
Best Local Similarity 26.1%; Pred. No. 11;
Matches 65; Conservative 35; Mismatches 84; Indels 65; Gaps 14;

Qy 8 VTERRSVHFIAEALTENCRIE-ERRRHVLVGI-----SPFNSRFSSEDIYIRLIGW 57
Db 83 VIESKDLKFIENKL-----RELLDEYKRVKIGKLGYPDPLETPPSIYEINKVIGG 138
Qy 58 AKAQFQSVSVLL-----AGHEAANLLEALGTGPRGKAERKVRKESVNRN 99
Db 139 YEADTRMIEVLYTNTVYKEYLKAKDKICSDVARYLLE-----GKGKIDKSSAN 192
Qy 100 RFAERLVAHGGDPKAIHTFSDPIDNKAYQLLRQVEHAFPEQHERHACLMSREALI 159
Db 193 ----VSMVLVNYG--VTSVQKLDFIDINSVGVRSKKET-----KGILLD---EALA 236
Qy 160 GRARGVSLMMEEVSEDMNLAV---BYVIAELPFFIGAPDILEVEE-TLLAYHRPWKLG 215
```



```
Db 237 GNAL-VSEYRTINEEYLSAIVKADYIKNNLQHERGFRDIKEIDNTITKVPYLEP----- 290
QY 216 KISNHEPSI 224
Db 291 -ISNSEASI 298

RESULT 14
O75130 PRELIMINARY; PRT; 864 AA.
AC O75130
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE KIAA0635 protein (Fragment).
GN Name=KIAA0635;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL; AB014535; BAA31610.2; -.
FT NON TER 1
SQ SEQUENCE 864 AA; 101410 MW; 19497C0E23095151 CRC64;

Query Match 7.58; Score 97; DB 2; Length 864;
Best Local Similarity 21.08; Pred. No. 35;
Matches 59; Conservative 56; Mismatches 104; Indels 62; Gaps 12;

QY 2 NEMTGMVTERRSVFIAEALTENCREFERRRHVLVGISPFNS-----RFSEDIYVRL- 54
DB 549 DDLATMARENQETISLELEAAVQEKMSRVKHYITEVSWESLMAAKEKQDLDLRFQ 608
QY 55 -----IGWAKAQKSVSVLLAGHEAANLEALGTPGKAERKVRKEVSNRRPFAE 104
DB 609 MLHNRADWEVKAHQAGESSVRL-----LLSIDTER---RHLRERVELLEKEIQ 657
QY 105 RALVAHGDDPKAHTTFSDFDN--KAYQLLQEVHEHAFPPHPRHACLMS--REALIG 160
DB 658 EHINAH-----HAYESQISSMAKAMSRLEELRHOEDEKATVLN---DLSSRLRLCIK 707
QY 161 RARGVSLMVEESVSDMLNLAVEYVIAELPPFIGAPDILEVE-----ETLLAYHR 209
DB 708 LDSGKDITMQLNSK--NLSEFVVVLELVNKSSEDLKQLNSRHTVKNLESLLATNR 765
QY 210 PWKLGEKISNHFCSIRPNQGYLIVQEMAMLSEKRITSE 250
DB 766 DKPFHSLTSHS-----KDTEIQLLKEKIT--LSESKLTSQ 799

RESULT 15
BRC2 HUMAN
ID BRC2 HUMAN STANDARD; PRT; 3418 AA.
AC P51587; O01813; O15008; Q13879;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Breast cancer type 2 susceptibility protein.
GN Name=BRC2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```

RN RP SEQUENCE FROM N.A.
RX MEDLINE=96112016; PubMed=8524414;
RA Wooster R., Bignell G., Lancaster J., Swift S., Seal S., Mangion J.,
RA Collins N., Gregory S., Gumbs C., Micklem G., Barfoot R., Hamoudi R.,
RA Patel S., Rice C., Biggs P., Hashim Y., Smith A., Connor F., Tonin P.,
RA Arason A., Gudmundsson J., Ficenec D., Kelsell D., Ford D., Tonin P.,
RA Bishop D.T., Spurr N.K., Ponder B.A.J., Seles R., Peto J., Devilee P.,
RA Corneliisse C., Lynch H., Narod S., Lenoir G., Egilsson V.,
RA Barkadottir R.B., Easton D.F., Bentley D.R., Futreal P.A.,
RA Ashworth A., Stratton M.R.;
RT "Identification of the breast cancer susceptibility gene BRCA2.";
RL Nature 378:789-792(1995).
RN [2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=96172838; PubMed=8589730;
RA Tavtigian S.V., Simard J., Rommens J., Couch F., Shattuck-Bidens D.,
RA Neuhausen S., Merajver S., Thorlacius S., Offit K., Stoppa-Lyonnet D.,
RA Belanger C., Bell R., Berry S., Bogden R., Chen Q., Davis T.,
RA Dumont M., Frye C., Hattier T., Jammulapati S., Janekci T., Jiang P.,
RA Kahrer R., Leblanc J.-F., Mitchell J.T., McArthur-Morrison J.,
RA Nguyen K., Peng Y., Samson C., Schroeder M., Snyder S.C., Steele L.,
RA Stringfellow M., Stroup C., Swedlund B., Swensen J., Teng D.,
RA Thomas A., Tran T., Tran T., Tranchant M., Weaver-Feldhaus J.,
RA Wong A.K.C., Shizuya H., Eyfjord J.B., Cannon-Albright L., Labrie F.,
RA Skolnick M.H., Weber B., Kamb A., Goldar D.E.;
RT "The complete BRCA2 gene and mutations in chromosome 13q-linked
RT kindreds.";
RL Nat. Genet. 12:333-337(1996).
RN [3]
RN RP SEQUENCE FROM N.A., AND VARIANTS HIS-289; GLN-322; ASN-372; VAL-784;
RX SER-929; PHE-976; ILE-987; ASP-991; ASN-1561; LYS-1880; MET-1915;
RP PHE-2138; ARG-2162; ARG-2440; VAL-2466; THR-2490; PRO-2835; ALA-2856;
RP PHE-2944; THR-2951; ILE-3244 AND VAL-3412.
RA Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.,
RT "NIHNS-SNPs, environmental genome project, NIHNS ES15478, Department
RT of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu)";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RN RP SEQUENCE FROM N.A.
RX PubMed=15057823; DOI=10.1038/nature02379;
RA Dunham A., Matthews L.H., Burton J., Ashurst J.L., Howe K.L.,
RA Ashcroft K.J., Beare D.M., Burford D.C., Hunt S.E.,
RA Griffiths-Jones S., Jones M.C., Keenan S.J., Oliver K., Scott C.E.,
RA Ainscough R., Almeida J.P., Ambrose K.D., Andrews D.T.,
RA Ashwell R.I.S., Babbage A.K., Baggeley C.L., Bailey J., Bannerjee R.,
RA Brown A.J., Brown J.Y., Burrill W., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M.E., Clark S.Y., Clarke G., Clee C.M.,
RA Clegg S.C., Cobley V., Collins J.E., Corby N., Coville G.J.,
RA Deloukas P., Dhami P., Dunham I., Dunn M., Earthrowl M.E.,
RA Ellington A.G., Faulkner L., Frankish A.G., Frankland J., French L.,
RA Garner P., Garnett J., Gilbert J.G.R., Gilson C.J., Ghori J.,
RA Grafham D.V., Gribble S.M., Griffiths C., Hall R.E., Hammond S.,
RA Hartley J.L., Hart E.A., Heath P.D., Howden P.J., Huckle E.J.,
RA Hunt P.J., Hunt A.R., Johnson C., Johnson D., Kay M., Kimberley A.M.,
RA King A., Laird G.K., Langford C.J., Lawlor S., Leongamornlert D.A.,
RA Lloyd D.M., Lloyd C., Loveland J.E., Lovell J., Martin S.,
RA Mashreghi-Mohammadi M., McLaren S.J., McMurray A., Milne S.,
RA Moore M.J.F., Nickerson T., Palmer S.A., Pearce A.V., Peck A.I.,
RA Pelan S., Phillimore B., Porter K.M., Rice C.M., Searle S.,
RA Sehra H.K., Showkeen R., Skuce C.D., Smith M., Steward C.A.,
RA Sycamore N., Tester J., Thomas D.W., Tracey A., Tromans A., Tubby B.,
RA Wall M., Wallis J.M., West A.P., Whitehead S.L., Willey D.L.,
RA Wilming L., Wray P.W., Wright M.W., Young L., Coulson A., Durbin R.,
RA Hubbard T., Sulston J.E., Beck S., Bentley D.R., Rogers J., Ross M.T.;
RT "The DNA sequence and analysis of human chromosome 13.";
RL Nature 428:522-528(2004).
RN [5]
RN RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 1519-1551 IN COMPLEX WITH
RP RAD51.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 12:47:24 ; Search time 155.579 Seconds
(without alignments)
1497.801 Million cell updates/sec

Title: US-10-627-124-2

Perfect score: 2080

Sequence: 1 MSOSIKLFSVLSPQNNFY.....AESGLYTRGPVSLVAFDGA 405

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2080	100.0	405	1 CYPX_BACSU	Q34926 bacillus su
2	958	46.1	407	2 Q7N9M6	Q7n9m6 photorhabdu
3	825.5	39.7	408	2 Q83VE9	Q83ve9 streptomyc
4	788.5	37.9	407	2 Q9RJO7	Q9rjo7 streptomyc
5	497.5	23.9	404	2 Q73WL1	Q73wl1 mycobacteri
6	497.5	23.9	404	2 AAS04966	Aas04966 mycobacte
7	468	22.5	405	1 CPXK_SACER	P33271 saccharopol
8	455.5	21.9	412	2 Q70AS7	Q70as7 streptomyc
9	455.5	21.9	412	2 CAE53704	Caes53704 streptom
10	444	21.3	404	2 Q81Q07	Q81q07 bacillus an
11	444	21.3	404	2 AAT31748	Aat31748 bacillus
12	442	21.2	404	2 Q6H186	Q6h186 bacillus th
13	438	21.1	404	2 Q73719	Q73719 bacillus ce
14	438	21.1	404	2 AAS41573	Aas41573 bacillus
15	433.5	20.8	404	2 Q83X67	Q83x67 streptomyc
16	429.5	20.6	405	2 Q82E54	Q82e54 streptomyc
17	427	20.5	411	2 Q74437	Q74437 mycobacteri
18	427	20.5	411	2 AAS02661	Aas02661 mycobacte
19	424.5	20.4	400	1 C13C_XYLFA	Q9p966 xyella fas
20	424.5	20.4	424	2 Q93Q05	Q93q05 pseudomonas
21	422.5	20.3	424	2 Q7BRJ9	Q7brj9 pseudomonas
22	422.5	20.3	424	2 AAR83738	Aar83738 pseudomon
23	418.5	20.1	396	1 YJIB_BACSU	Q34374 bacillus su
24	417.5	20.1	410	2 Q737F3	Q737f3 bacillus ce
25	417.5	20.1	410	2 AAS41609	Aas41609 bacillus
26	417	20.0	410	2 Q9X5P9	Q9x5p9 streptomyc
27	415.5	20.0	397	2 Q59523	Q59523 micromonosop
28	415.5	20.0	405	1 CPXM_BACSU	P27632 bacillus su
29	415	20.0	414	2 Q70AK9	Q70ak9 streptomyc
30	415	20.0	414	2 CAE53712	Caes53712 streptom
31	414.5	19.9	399	1 C13C_XYLFT	Q87av9 xyella fas

Q89g11 bradyrhizob
Q735a2 bacillus ce
Aas42160 bacillus
Q70j22 bacillus am
Caell256 bacillus
Q7nj97 gloeobacter
Q6089 bacillus me
P77902 mycobacteri
Q73xe4 mycobacteri
Aas04682 mycobacte
Q59910 streptomyc
Q87ax5 xyella fas
Q08459 bacillus su
Q9f2q0 streptomyc

ALIGNMENTS

RESULT 1
CYPX_BACSU STANDARD; PRT; 405 AA.
AC Q34926;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Putative cytochrome P450 CYPX (EC 1.14.-.-).
GN Name=cypX; Synonyms=cypB, cyp134; OrderedLocusName=BSU35060;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Soldo B., Rivolta C., Reynolds S., Maue C.,
RA Karamata D.;
RT "Nucleotide sequence of the 300-304 chromosomal segment of Bacillus
RT subtilis."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=38044033; PubMed=9384377; DOI=10.1038/36786;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conner J.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiteh J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidis A., Lardinois S., Lauber J.,
RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C.,
RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Presecan E., Fujic P., Furnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P.,
RA Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,
RA Takenaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";

405

```

RL Nature 390:249-256(1997).
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AF017113; AAC67280.1; -.
DR EMBL; Z99121; CAB15511.1; -.
DR PIR; F69611; F69611.
DR HSP; P33006; 1CPT.
DR Subtilisin; BGI2580; cypX.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 2.
DR Pfam; PR00359; BP450.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Complete proteome; Heme; Hypothetical protein; Monooxygenase;
KW Oxidoreductase.
FT METAL 353 353 Iron (heme axial ligand) (By similarity).
SQ SEQUENCE 405 AA; 45473 MW; A943DDFEAB67BD01 CRC64;

Query Match 100.0%; Score 2080; DB 1; Length 405;
Best Local Similarity 100.0%; Pred. No. 1.1e-141;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSQSIKLFVSLDQFNPNYPAYSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 60
Db 1 MSQSIKLFVSLDQFNPNYPAYSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFT 60
QY 61 TKSLVERAEPVMRGPVLAQMGHKEHSAKRIVVRSFTGDALDHLSPLIKQNAENLLAPYL 120
Db 61 TKSLVERAEPVMRGPVLAQMGHKEHSAKRIVVRSFTGDALDHLSPLIKQNAENLLAPYL 120
QY 121 ERGKSLVNDFGKTFACVVTMDMLGDKRDHEKISEHSGVADFTISQSPEARASLW 180
Db 121 ERGKSLVNDFGKTFACVVTMDMLGDKRDHEKISEHSGVADFTISQSPEARASLW 180
QY 181 CSQQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSKDILALINLVLLAATPADK 240
Db 181 CSQQLSQYLMPIKERRVNPVSGDLISILCTSEYEGMALSKDILALINLVLLAATPADK 240
QY 241 TLALMYHLNNPEQNDVLAADSLVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKK 300
Db 241 TLALMYHLNNPEQNDVLAADSLVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKK 300
QY 301 DTIVFCMIGANRDPEAFQDPVFNHREDLGKSAFSGAARHLAFSGIHNCVGAFAK 360
Db 301 DTIVFCMIGANRDPEAFQDPVFNHREDLGKSAFSGAARHLAFSGIHNCVGAFAK 360
QY 361 NEIEIVANIVLDKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405
Db 361 NEIEIVANIVLDKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405

RESULT 2
Q7N9M6 PRELIMINARY; PRT; 407 AA.
ID Q7N9M6
AC Q7N9M6
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Similar to cytochrome P450.
GN OrderedLocusNames=plu0296;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=TT01;
RC MEDLINE=22957627; PubMed=14528314;
RX Duchaud E., Runicok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taouit S., Bocs S., Bouraux-Bude C., Chandelier M., Charles J.-F.,
RA Dassa E., Derose R., Derselle S., Freyssinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Sigquier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Bomare N., Danchin A., Kunst F.,
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313(2003).
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; BX571859; CAE12591.1; -.
DR Photolista; plu0296; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Complete proteome; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 407 AA; 46297 MW; 7C3F545F86538D65 CRC64;

Query Match 46.1%; Score 958; DB 2; Length 407;
Best Local Similarity 46.6%; Pred. No. 9.2e-61;
Matches 185; Conservative 82; Mismatches 124; Indels 6; Gaps 3;

QY 8 FSVLSDFQFNPNYPAYSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFTTKSILVER 67
Db 7 FNIHDFKPKINPYDFDILHKQDLVYFEQSQNSYFQKYEDVDAILK-SSIFNTKPLTAL 65
QY 68 AEPVWRGVPVLAQMGHKEHSAKRIVVRSFTGDALD-HLSPLIKQNAENLLAPYLERGKSD 126
Db 66 AEPVMDRVLQAQMEGHEHACKRFIMQGLSRDVFNRYEPMIRKITEDLIQPYMEKGNID 125
QY 127 LVNDFKTFACVVTMDMLGDKRDHEKISEHSGVADFTISQSPEARASLWSCSEQLS 186
Db 126 LVNDFORDVAVLTLSILGLPSDNYRDIAEWHKGIASFQFDQTELEKNSLECSQKLI 185
QY 187 QYLMPIKERRVNPVSGDLISILCTSEYEGMALSKDILALINLVLLAATPADKTLALMI 246
Db 186 RLLKPIIDQRRNPSKDIISIFC----QDTAMSMSEITALCLINLLAATEPADKILAMML 241
QY 247 YHLLNPEQNDVLAADSLVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKKDTIVFC 306
Db 242 NHLISNPSMLDVVLKDRSLVRDAFEETRLTSPVQLIPREASEDVTISGIDIPKGAUVFC 301
QY 307 MIGAANRDPEAFQDPVFNHREDLGKSAFSGAARHLAFSGIHNCVGAFAKNEIEIV 366
Db 302 MIGAANRDPEAFQDPVFNHREDLGKSAFSGAARHLAFSGIHNCVGAFAKNEIEIV 361
QY 367 ANIVLDMKRNIRLEEDFCYAESGLYTRGPVSLVAFD 403
Db 362 SNIIIDLLHNLPAADHYHYQETGVYTRGPKSLLSFD 398

RESULT 3
Q83YE9 PRELIMINARY; PRT; 408 AA.
ID Q83YE9
AC Q83YE9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Shy1.
GN Names=shy1;
OS Streptomyces hygroscopicus subsp. yingchengensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=228732;
RN [1]
RP SEQUENCE FROM N.A.

```

```
RA Qin L., Wang H., Wu Z., Lu Y.T., Deng Z., Zhao G.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AY260760; AAP21648.1; -.
DR HSP; P33006; ICPT.
DR GO; GO:0004497; F:monoxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
KW Heme; Monoxygenase; Oxidoreductase.
SQ SSQUENCE 408 AA; 44826 MW; B7137737217CDE05 CRC64;

Query Match 39.7%; Score 825.5; DB 2; Length 408;
Best Local Similarity 42.8%; Pred. No. 3.3e-51;
Matches 173; Conservative 75; Mismatches 153; Indels 3; Gaps 3;

QY 1 MSQSIKLFVLSQFQNNPYAYFSQLEEDPVHYEIDSFYFSRYHDV-RYLQHPDI-F 59
DB 1 MSAAHLPDILSPFAANPYAVAVMREKEPLIWHTEATQSYIISRYEDVERVFKKAEF 60
QY -60 TTSLVERAEPVNRGVPVLAQMKGHSKRIIVRSFIDALDH-LSPLIKQNAENLLAP 118
DB 61 TTDYNNWQLEPV-HGRTILQSGREHARRALVAPAFRGDLQKFLFVIERNRELIDA 119
QY 119 YLGRGSDLVNDGFKTFAVCVTWMLGDKRDHEKISEHSGVADFTITSQSPARAH 178
DB 120 FRHTGSADIVNDYATRPVNVNVIADMLGLDKADHARFGHTVATVAFIGNLSGDEVAAG 179
QY 179 LMCSEQLSQYLMPIKERRVNPVNSDLSILCTSEYEGMALSKDILALINVLAAPEA 238
DB 180 ERTVEFAEYMLPVIRERRANPGDILLSTLCAAEVDGVRMSDEIKAFCSLLAAGETT 239
QY 239 DKTALMIYHLNPNPEQMDVADRSLVPRAIATRYKPPVQLIPRLSQDVTVGMEI 298
DB 240 DKAIGILANLLSHPDQLAARVDRSLIPAAFAETLRYTPPVQIMRQSATDVEVTGTTI 299
QY 299 KDTIVFCMIGANRDPEAFQPDVFNHREDIGKSAFSAARHAFSGIHCNVCVAAF 358
DB 300 PAGATVCLIGANRDERRYDRDFIDFDDLTATTSFAAAGHLAFALGRHFCVGALL 359
QY 359 AKNEIEIVANIVLDKRNIRLEEDFCVAESGLYTRGPVSLVAF 402
DB 360 AKAEVGLNQLDAMPDLALDGHDLVEQGVETRGKTLPVRF 403

RESULT 4
Q9RJQ7 PRELIMINARY; PRT; 407 AA.
AC Q9RJQ7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative cytochrome P450.
GN OrderedLocuNames=SCO0583; ORFNames=SCF55.07;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.B., Quail M.A., Kieser H.,
RA Harper D., Batenan A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,

Hopwood D.A.;
*Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AL939106; CAB61277.1; -.
DR HSP; P33006; ICPT.
DR GO; GO:0004497; F:monoxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002397; BP450.
DR Pfam; PF00067; p450; 2.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
KW Complete proteome; Heme; Monoxygenase; Oxidoreductase.
SQ SSQUENCE 407 AA; 44916 MW; 86320B142522C788 CRC64;

Query Match 37.9%; Score 788.5; DB 2; Length 407;
Best Local Similarity 41.5%; Pred. No. 1.5e-48;
Matches 169; Conservative 73; Mismatches 162; Indels 3; Gaps 3;

QY 1 MSQSIKLFVLSQFQNNPYAYFSQLEEDPVHYEIDSFYFSRYHDV-RYLQHPDIF 59
DB 1 MSTAQQVPDILSPFAANPYATRWDSAPLIRHEATQSWIYSRYEDVERVFKDRAGQF 60
QY 60 TTSLVERAEPVNRGVPVLAQMKGHSKRIIVRSFIFG-DALDHLSPLIKQNAENLLAP 118
DB 61 TTDYNNWQLEPV-HGRTILQSGREHARRALVAPAFRGDLQKFLFVIERNRELIDA 119
QY 119 YLGRGSDLVNDGFKTFAVCVTWMLGDKRDHEKISEHSGVADFTITSQSPARAH 178
DB 120 FRHTGSADIVNDYATRPVNVNVIADMLGLDKADHARFGHTVATVAFIGNLSGDEVAAG 179
QY 179 LMCSEQLSQYLMPIKERRVNPVNSDLSILCTSEYEGMALSKDILALINVLAAPEA 238
DB 180 ARTTFFAEYMLPVIRERRANPGDILLSTLCAAEVDGVRMSDEIKAFCSLLAAGETT 239
QY 239 DKTALMIYHLNPNPEQMDVADRSLVPRAIATRYKPPVQLIPRLSQDVTVGMEI 298
DB 240 DKAIGILANLLSHPDQLAARVDRSLIPAAFAETLRYTPPVQIMRQSATDVEVTGTTI 299
QY 299 KDTIVFCMIGANRDPEAFQPDVFNHREDIGKSAFSAARHAFSGIHCNVCVAAF 358
DB 300 PAGATVCLIGANRDERRYDRDFIDFDDLTATTSFAAAGHLAFALGRHFCVGALL 359
QY 359 AKNEIEIVANIVLDKRNIRLEEDFCVAESGLYTRGPVSLVAFDGA 405
DB 360 AKAEVGLNQLDAMPDLALDGHDLVEQGVETRGKTLPVRF 406

RESULT 5
Q73WL1 PRELIMINARY; PRT; 404 AA.
AC Q73WL1
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocuNames=MAP2649c;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OX NCBI_TaxID=1770;
RN [1]
RC STRAIN=N.A.
RP SEQUENCE FROM N.A.
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AE017236; AAS04966.1; -.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
```

```
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN 1.
KW Complete proteome; Heme; Hypothetical protein; Monooxygenase;
SQ Oxidoreductase.
KW SEQUENCE 404 AA; 45054 MW; 3224F829F78A27FE CRC64;

Query Match 23.9%; Score 497.5; DB 2; Length 404;
Best Local Similarity 29.8%; Pred. No. 1.5e-27;
Matches 119; Conservative 86; Mismatches 161; Indels 33; Gaps 7;

QY 18 NPAYFSQLEEDPVH-----YESIDSYFISRVHVDYIQLHPDIFTTKSLVERAE 69
DB 24 DPYELARLRREPIQLRDTSGMPHEESKPVFIVYRHEDAQQLMRDNETSSAAVIAFG 83
QY 70 PVMRGPVLAQMGKHSKRIVVRISFGDALDHL-SPLIKQNAENLLAPYLRGKSDLV 128
DB 84 PVLGERVMLGIDEPVHGRLSLVSKAFSQKALARWEDELVGRVGNLSLIDRFAGNGKADLV 143
QY 129 NDFGKTFACVVTMDLGLDKRDHEKISEHSGVADFTITSQSPEARAHSLWCSEOLSOY 188
DB 144 KEFTDPSPRIIAGLGLPEQDPFPQQRWISILLSWIL-----NPE---RGLAASAAALCDY 196
QY 189 LMPVIKERRVNPGSDLISILCTSEYEGMALSDKDIALLINVLAAATEPADKTLALMIYH 248
DB 197 FAPILARRAEAPKDDLLISGLAQAEIDGKLEDEEIIYFLLRLLPAGVETTYRALSILLA 256
QY 249 LLNPNQMDVLAQRSVPRATETLYRKPVPQLIPRQLSQDVTGVGMEIKKDTIVFCMI 308
DB 257 LLSDPQLDAIRGDSLLPQAEIEGVWRWEPPLLTITRVATRDTELGGVPIPGAGSTVMPML 316
QY 309 GAANROPEAFEPQDVNFIHREDLGKISAFSGAARHLAFSGIHCVCVGAFAAKNEIEIVAN 368
DB 317 GAANROEDRYPPDRFDI-----FRAPKSHLGWGHVCLGMHLARLEMTAVN 366
QY 369 IVLDMKNRINRLEEDFCYAES----GLYTRGPVSLLVAFD 403
DB 367 LLLDRLPNRLDPD---ADDPHIRGQVFRSPTSPVPLVD 402

RESULT 7
CPXK_SACER STANDARD; PRT; 405 AA.
ID CPXK_SACER
AC P33271;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Cytochrome P450 107B1 (EC 1.14.-.-) (P450CYP107B1).
GN Names=CYP107B1;
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 172-203 AND 302-321.
RC STRAIN=NRRL 2338;
RX MEDLINE=92121109; PubMed=1732208;
RA Andersen J.F., Hutchinson C.R.;
RT "Characterization of Saccharopolyspora erythraea cytochrome P-450
RL J. Bacteriol. 174:725-735(1992).
CC -I- FUNCTION: Not known, probably involved in the catabolism of octane
CC and guaiacol. It displays a weak activity in the O-dealkylation of
CC 7-ethoxycoumarin.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR ENBL; M83110; AAA26483.1; -.
DR PIR; B42606; B42606.
DR HSP; Q00441; 10XA.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 2.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Direct protein sequencing; Electron transport; Heme; Membrane;
KW Monooxygenase; Oxidoreductase.
FT METAL 352 352
SQ SEQUENCE 405 AA; 45238 MW; 71C93CEC1FDC53FD CRC64;

Query Match 22.5%; Score 468; DB 1; Length 405;
Best Local Similarity 30.0%; Pred. No. 2e-25;
Matches 115; Conservative 85; Mismatches 149; Indels 34; Gaps 8;
```



```
QY 13 DOFQNNPYAFSOLREE--DPVHYEBSIDSFYISRYHDVRYILOHP-----DIFTTKSLVE 66
Db 14 DAFQDRHNRKARMEEPQRIRTVNGLDLAWLITRYEDVKQALLDPRIAKDFGRTOQIIE 73
QY 67 R--AEPVMR-----GPVLAQMHGKHSKRIIVRSFGDGLDHLSPLIKQNAENLLA 117
Db 74 KRLADARRPGFSDIGPHMLNTDPDHTRLKLVKAFARVEGLRPRIEQITDDLLD 133
QY 118 PYLERGKSDLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFPITISISPEARAH 177
Db 134 RLAGESEVLDLIDFAFPLPITVISELGMVEDSRDDFRSNTWNLVD-----GSQPEAAQ 188
QY 178 SLWCSEQLSOYLMMPVIKERRVNPFGSOLISILCTSEYEGMALSKOILALINVLAAATEP 237
Db 189 A---SVAMVEYLTELIAKRTKTEPGDOLLTALLEAVEDGDRLSGELIAMVFLLLVAGHET 245
QY 238 ADKTLALMIYHLNPEQMDVLADRSILVPRATAETLYKPPV-QLIPQLSODTVVGGM 296
Db 246 TYNLGNVLSLGNPDQALARNPDLPSLLPGAIEETLRYESPANGTFRHTAEAVRFGDV 305
QY 297 EIKKDTIVFCMGAANRDPFAEPQDPVFNHREDLGIKSAFSGAARHLAFSGIHNCVGA 356
Db 306 VIEGELVWALGAANDGERFEDPREDITRTTG-----HVAEGHGHFCVGA 355
QY 357 AFAKNEIEIVANVLDKRNIRL 379
Db 356 ALARLEAQIAGVRLERFPDLRM 378
RESULT 8
Q70AS7
ID Q70AS7 PRELIMINARY; PRT; 412 AA.
AC Q70AS7
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Putative cytochrome P450 reductase (EC 1.14.14.1).
GN Name=CYP0549;
OS Streptomyces peucetius;
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1950;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27952;
RA Parajuli N., Basnet D.B., Lee H.C., Sohng J.K., Liou K.;
RT "Genome analyses of Streptomyces peucetius ATCC 27952 for the
RT identification and comparison of cytochrome P450 complement with other
RT Streptomyces.";
RL Arch. Biochem. Biophys. 425:233-241(2004).
DR EMBL; A605536; CA53704.1; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR InterPro; IPR02397; BP450.
DR IntraPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
SQ SEQUENCE 412 AA; 45791 MW; 3F0D2529582C4F6A CRC64;
Query Match 21.9%; Score 455.5; DB 2; Length 412;
Best Local Similarity 29.3%; Pred. No. 1.7e-24;
Matches 125; Conservative 81; Mismatches 180; Indels 41; Gaps 12;
QY 1 MSQSIKFSVLSQFQNNPYAFSOLREEEDPVHYEBSIDSFYISRYHDVRYILOHPDIFT 60
Db 1 MTQSL-LHQILDYANRADPPIVEELR-KTPVHHEED-GPYVVGTYEIRSLLDHPRISS 57
QY 61 -TKSLVERA-EPVWRG-----PVLQMHGKHSKRIIVRSFGDGLDHLSPLIK 109
Db 58 DARNLASTAGDPLAESAEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
QY 110 QNAENLLAPYL-----ERGKSDLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFI 165
Db 118 GEHUGSVGLIDIGDGTGRIDLVESQYPPVTVICLLGVPREDEARFTWADTIA--- 174
QY 166 TSISQSPA-----RAHSLWCSEQLSOYLMMPVIKERRVNPFGSOLISILCTSEYEGMAL 218
Db 175 ASLDPNFADPAERKASHD--ARTQLGMVLAGLIBERRKNPGDDILSELATAKRGDGTW 232
QY 219 SKDILALINVLAAATEPADKTLALMIYHLNPEQMDVLADRSILVPRATAETLYKPP 278
Db 58 DARNLASTAGDPLAESAEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
Query Match 21.9%; Score 455.5; DB 2; Length 412;
Best Local Similarity 29.3%; Pred. No. 1.7e-24;
Matches 125; Conservative 81; Mismatches 180; Indels 41; Gaps 12;
QY 1 MSQSIKFSVLSQFQNNPYAFSOLREEEDPVHYEBSIDSFYISRYHDVRYILOHPDIFT 60
Db 1 MTQSL-LHQILDYANRADPPIVEELR-KTPVHHEED-GPYVVGTYEIRSLLDHPRISS 57
QY 61 -TKSLVERA-EPVWRG-----PVLQMHGKHSKRIIVRSFGDGLDHLSPLIK 109
Db 58 DARNLASTAGDPLAESAEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
```

```
QY 110 QNAENLLAPYL-----ERGKSDLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFI 165
Db 118 GEHUGSVGLIDIGDGTGRIDLVESQYPPVTVICLLGVPREDEARFTWADTIA--- 174
QY 166 TSISQSPA-----RAHSLWCSEQLSOYLMMPVIKERRVNPFGSOLISILCTSEYEGMAL 218
Db 175 ASLDPNFADPAERKASHD--ARTQLGMVLAGLIBERRKNPGDDILSELATAKRGDGTW 232
QY 219 SKDILALINVLAAATEPADKTLALMIYHLNPEQMDVLADRSILVPRATAETLYKPP 278
Db 233 TTMELLSTAALLIAGHETTIVNLVTNGMLTLRLNPVQLRQRLADPRLAVPIVEELLARFEP 292
QY 279 PVOLIPQLS-QPTVVGGMEIKDITVFCMIGAANRDPFAEPQDPVFNHREDLGIKSAF 337
Db 293 PVQLVQRTTLTDIEVRGVTIPKGAASLWLVLASGNRDPQRFEDRFPDPRDI----- 346
QY 338 SGAARHLAFSGIHNCVGAFAKNEIEIVANVLDKRNIRLEEDFCYAESGLYTRGPVS 397
Db 347 ----QHGLGSGIHSCFGAPLARLEAHIALSELARLENPRLLLEDPPIYKQNAVLRGPRH 402
QY 398 LLVAFDG 404
Db 403 LPIACDG 409
RESULT 9
CAE53704
ID CAE53704 PRELIMINARY; PRT; 412 AA.
AC CAE53704;
DT 20-MAY-2004 (T-EMBLrel. 27, Created)
DT 20-MAY-2004 (T-EMBLrel. 27, Last sequence update)
DT 20-MAY-2004 (T-EMBLrel. 27, Last annotation update)
DE Putative cytochrome P450 reductase (EC 1.14.14.1).
GN CYP0549.
OS Streptomyces peucetius;
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1950;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27952;
RA Parajuli N., Basnet D.B., Lee H.C., Sohng J.K., Liou K.;
RT "Genome analyses of Streptomyces peucetius ATCC 27952 for the
RT identification and comparison of cytochrome P450 complement with other
RT Streptomyces.";
RL Arch. Biochem. Biophys. 425:233-241(2004).
DR EMBL; A605536; CAE53704.1; -.
KW Oxidoreductase.
SQ SEQUENCE 412 AA; 45791 MW; 3F0D2529582C4F6A CRC64;
Query Match 21.9%; Score 455.5; DB 2; Length 412;
Best Local Similarity 29.3%; Pred. No. 1.7e-24;
Matches 125; Conservative 81; Mismatches 180; Indels 41; Gaps 12;
QY 1 MSQSIKFSVLSQFQNNPYAFSOLREEEDPVHYEBSIDSFYISRYHDVRYILOHPDIFT 60
Db 1 MTQSL-LHQILDYANRADPPIVEELR-KTPVHHEED-GPYVVGTYEIRSLLDHPRISS 57
QY 61 -TKSLVERA-EPVWRG-----PVLQMHGKHSKRIIVRSFGDGLDHLSPLIK 109
Db 58 DARNLASTAGDPLAESAEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
QY 110 QNAENLLAPYL-----ERGKSDLVNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFI 165
Db 118 GEHUGSVGLIDIGDGTGRIDLVESQYPPVTVICLLGVPREDEARFTWADTIA--- 174
QY 166 TSISQSPA-----RAHSLWCSEQLSOYLMMPVIKERRVNPFGSOLISILCTSEYEGMAL 218
Db 175 ASLDPNFADPAERKASHD--ARTQLGMVLAGLIBERRKNPGDDILSELATAKRGDGTW 232
QY 219 SKDILALINVLAAATEPADKTLALMIYHLNPEQMDVLADRSILVPRATAETLYKPP 278
Db 58 DARNLASTAGDPLAESAEEDSALPFSFLRLDPPHDLRMRNRSFGPPHSPHRVDGMR 117
```

```

Db 233 TTMELLSTAALLIAGHETTIVNLVTNGMLTLRLNPDVLQRLADPRLAVPIVEBLLRFEP 292
QY 279 PVOLIPQLS-QDTVVGMKIKDITVFCMIGANRDPEAFQPDVFNHIREDLGIKSAF 337
Db 293 PVOLIPQRTTLTDIEVGVTPIPKASLWULVLSGNRPQPFDPDRPDPRDI----- 346
QY 338 SGAARLAFSGIHNCVGAFAFNAKNEITEIVANIVLDKMRNIRLEEDFCYABSGLYTRGPVS 397
Db 347 ----QHLGLSGIHSCEGAPLARLEAHIALSELARLENRLLEDPPPYRQNAVILGPRH 402
QY 398 LVAFDG 404
Db 403 LPIACDG 409

RESULT 10
ID Q81Q07 PRELIMINARY; PRT; 404 AA.
AC Q81Q07; Q6HV76; Q6KS91;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Cytochrome P450 family protein.
GN OrderedLocusNames=BA2632, BAS2452; ORFNames=GBAA2632;
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;

[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Porton;
RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;
RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Ostad O.A., Helgason E., Rinkstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA Salzberg S.L., Thomson B., Friedlander A.M., Kohler T.M.,
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.;
RT "Bacillus anthracis comparative genomics."
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=Sterne;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AE017032; AAP26480.1; -
DR EMBL; AE017334; AAT31748.1; -
DR EMBL; AE017225; AAT54763.1; -
DR HSSP; Q55080; 1107.
DR TIGR; BA2632; -
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 2.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00359; BP450.

```

```

DR PROSITE; PS00086; CYTOCHROME P450; UNKNOWN 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 404 AA; 46728 MW; 960552E1D7792F55 CRC64;

Query Match 21.3%; Score 444; DB 2; Length 404;
Best Local Similarity 30.0%; Pred. No. 1.le-23;
Matches 114; Conservative 68; Mismatches 176; Indels 22; Gaps 7;

QY 18 NPAYFSQLEEDPVHYEESIDSYFISRYHDVRYILQHPDIFTTKSIVERAE-----PVMR 73
Db 24 NPYEWYQFMRDNHPVHYDDQDVWNVFLYDDVNRVLSDSYLSFSSRR--ERRQFAIPPLET 81
QY 74 GPVLAQMHGKHSKRRIIVRSFIGDALDHLSPKLIKONAEMLAPYLERKCKSLVNDPFGK 133
Db 82 RININSTDPPEHRNVRISVSKAFTPRSLQWKPRIQSIANELVKDIENCSEVDIVEQFAA 141
QY 134 TFAVCVTMDMLGLDKRDHEKISEWHSGVADFITSIQSPEARAHSLWCSQLSOLMPVI 193
Db 142 PLPVTVISDLGVPVTTDRKKIKAW-SDILFMPYSKEKFNLDLDAEKGIALNFKAYLLPIV 200
QY 194 KERRVNPNGSGLISILCTSEYEGMALSDKDLAILNLVLLAATPADKTLALMIY-HLLNN 252
Db 201 QEKRYHLTDIIISDLIRAEYEGERLTDEEIVTFSGLLAAGNETTNNLIINSFYCFIADS 260
QY 253 PEQMDVLADRSIVPRAIAETLRYKPPVOLIIPQLSODTVVGGMEIKDITVFCMIGAN 312
Db 261 PATYKEVREKPKLISKAVEVLYRFPVTL-ARITEDTNIFGLMKKQDVWVWVSAAN 319
QY 313 RDEAFEPQDVFNHREDLGIKSAFGAARHLAGSGIHNCVGAFAKNIIEIVANIVLD 372
Db 320 LDEKFSQASKFNHR-----IGNEKHLTFGKPHFCGLGAPLARLEAEIALTTFIN 370
QY 373 KRNINPLEEDF-----CYAES 398
Db 371 APEKIALSPSFNIEQCILEN 390

RESULT 11
AAT31748 PRELIMINARY; PRT; 404 AA.
ID AAT31748 PRELIMINARY; PRT; 404 AA.
AC AAT31748;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE Cytochrome P450 family protein.
GN GBAA2632.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus Group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics."
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017334; AAT31748.1; -
SQ SEQUENCE 404 AA; 46728 MW; 960552E1D7792F55 CRC64;

Query Match 21.3%; Score 444; DB 2; Length 404;
Best Local Similarity 30.0%; Pred. No. 1.le-23;
Matches 114; Conservative 68; Mismatches 176; Indels 22; Gaps 7;

QY 18 NPAYFSQLEEDPVHYEESIDSYFISRYHDVRYILQHPDIFTTKSIVERAE-----PVMR 73
Db 24 NPYEWYQFMRDNHPVHYDDQDVWNVFLYDDVNRVLSDSYLSFSSRR--ERRQFAIPPLET 81
QY 74 GPVLAQMHGKHSKRRIIVRSFIGDALDHLSPKLIKONAEMLAPYLERKCKSLVNDPFGK 133
Db 82 RININSTDPPEHRNVRISVSKAFTPRSLQWKPRIQSIANELVKDIENCSEVDIVEQFAA 141
QY 134 TFAVCVTMDMLGLDKRDHEKISEWHSGVADFITSIQSPEARAHSLWCSQLSOLMPVI 193

```



```
Db 142 PLEVTWISDLGVPTTDRKKIKAW--SDILFMPYSKEKFNLDLDAEKGIALNEFKAYLLPIV 200
Qy 194 KERRVNPGLSILICTSEYEGMALSDKDLIALILNVLAAATPDKTALMIY-HLLNN 252
Db 201 QEKRYHLTDDIISDLIRAEYEGRLTDEEIVTSLGLLAAGNETTNNLIINSFYCFVLDS 260
Qy 253 PEQNDVLADRSIVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKDTIVFCMIGAA 312
Db 261 PATYKEVREKPKLISKAVEEVLRYRFPVTL-ARRITEDTNIIFGLMKDQWVAVWSAAN 319
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
RESULT 12
Q6HI86 PRELIMINARY; PRT; 404 AA.
ID Q6HI86
AC Q6HI86; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Cytochrome P450.
GN Name: cypA; ORFNames=BP9727.2414;
OS Bacillus thuringiensis serovar konkukian str. 97-27.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus thuringiensis serovar konkukian.
OX NCBI_TaxID=281309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=97-27;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AE017355; AA160007.1; -.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 404 AA; 46771 MW; C24B4C107972EFP4 CRC64;

Query Match 21.2%; Score 442; DB 2; Length 404;
Best Local Similarity 30.0%; Pred. No. 1.5e-23;
Matches 114; Conservative 68; Mismatches 176; Indels 22; Gaps 7;

Qy 18 NPAYFSQLEEDPVPVYEEISDSYFSRYHVDVRYLQHPDIFTTKSLVERAE-----PVMR 73
Db 24 NPVEYQFMEDNHPVHYDDEQVNVFLYDDVNRVLSYRLFSSRR--ERRQPAIPLET 81
Qy 74 GPVLAQMHGKHSKRIRVRSFPGDLDHLSPLIKQNAENLLAPYLERGKSLDNDFGK 133
Db 82 RININSTDPPEHNRVRSIVSKAFTPRSLQWKPRIOAIANELVQHICKYSEVNIIVEFAA 141
Qy 134 TFACVVTMDLGLDKRDHEKISEWHSGVADFTISQSPEARAHSLCSEQLSOYLMPIV 193
Db 142 PLPVTVISDLGVPTTDRKKIKAW--SDILFMPYSKEKFNLDLDAEKGIALNEFKAYLLPIV 200
Qy 194 KERRVNPGLSILICTSEYEGMALSDKDLIALILNVLAAATPDKTALMIY-HLLNN 252
Db 201 QEKRYHLTDDIISDLIRAEYEGRLTDEEIVTSLGLLAAGNETTNNLIINSFYCFVLDS 260
Qy 253 PEQNDVLADRSIVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKDTIVFCMIGAA 312
Db 261 PATYKEVREKPKLISKAVEEVLRYRFPVTL-ARRITEDTNIIFGLMKDQWVAVWSAAN 319
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
```

```
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
RESULT 13
Q73719 PRELIMINARY; PRT; 404 AA.
ID Q73719
AC Q73719; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Cytochrome P450 family protein.
GN OrderedLocusNames=BCE2659;
OS Bacillus cereus (Strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=22523;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14960714;
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT adaptations and a large plasmid related to Bacillus anthracis pXOI.";
RL Nucleic Acids Res. 32:977-988(2004).
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
DR EMBL; AE017272; AA541573.1; -.
DR TIGR; BCE2659; -.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Complete proteome; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 404 AA; 46830 MW; F0FD27F1CD865640 CRC64;

Query Match 21.1%; Score 438; DB 2; Length 404;
Best Local Similarity 29.5%; Pred. No. 2.9e-23;
Matches 112; Conservative 73; Mismatches 173; Indels 22; Gaps 7;

Qy 18 NPAYFSQLEEDPVPVYEEISDSYFSRYHVDVRYLQHPDIFTTKSLVERAE-----PVMR 73
Db 24 NPVEYQFMEDNHPVHYDDEQVNVFLYDDVNRVLSYRLFSSRR--ERRQFSIPLET 81
Qy 74 GPVLAQMHGKHSKRIRVRSFPGDLDHLSPLIKQNAENLLAPYLERGKSLDNDFGK 133
Db 82 RININSTDPPEHNRVRSIVSKAFTPRSLQWKPRIOAIANELVQHICKYSEVNIIVEFAA 141
Qy 134 TFACVVTMDLGLDKRDHEKISEWHSGVADFTISQSPEARAHSLCSEQLSOYLMPIV 193
Db 142 PLPVTVISDLGVPTTDRKKIKAW--SDILFMPYSKEKFNLDLDAEKGIALNEFKAYLLPIV 200
Qy 194 KERRVNPGLSILICTSEYEGMALSDKDLIALILNVLAAATPDKTALMIY-HLLNN 252
Db 201 QEKRYHLTDDIISDLIRAEYEGRLTDEEIVTSLGLLAAGNETTNNLIINSFYCFVLDS 260
Qy 253 PEQNDVLADRSIVPRAIAETLRYPVQLIPROLSDQTVVGGMEIKDTIVFCMIGAA 312
Db 261 PGTYKEUREEPTLISKAEIEVLRYPVTL-ARRITEDTNIIFGLMKDQWVAVWSAAN 319
Qy 313 RDEAFEPQDVFNHREDLGKSAFSGAARHLAFSGIHNVCVGAFAKNEIEIVANIVLD 372
Db 320 LDEKKFSQASKFNHR-----IGNEKHLTFGKGFHCLGAPLARLEAEIALTTFIN 370
Qy 373 KMRNIRLEEDF-----CYAES 388
Db 371 AFEKIALSPSFNIEQCILEN 390
```

```
Db 371 AFEKIALSPSFNLEQCILEN 390
RESULT 14
AAS41573
ID AAS41573 PRELIMINARY; PRT; 404 AA.
AC AAS41573;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE Cytochrome P450 family protein.
GN BCE2659.
OS Bacillus cereus (strain ATCC 10987).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=222523;
RN [1]
RP SEQUENCE FROM N.A.
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,
RA Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.P.,
RA Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.,
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic
RT adaptations and a large plasmid related to Bacillus anthracis pXOI.";
RL Nucleic Acids Res. 32:977-988(2004).
DR ENBL; AE017272; AAS41573.1; -.
DR TIGR; BCE2659; -.
SQ SEQUENCE 404 AA; 46830 MW; F0FD27F1CD865640 CRC64;
Query Match 21.1%; Score 438; DB 2; Length 404;
Best Local Similarity 29.5%; Pred. No. 2.9e-23;
Matches 112; Conservative 73; Mismatches 173; Indels 22; Gaps 7;
Qy 18 NPAYSQLREEDPVHYEESIDSYFISRVHVDVYILQHPDIFTKSLVERAE-----PVMR 73
Db 24 NPYEWQFMRDNHPVHYDEBQDVMNVFLYEDVARVLSYDLFSRR--ERRQFSIPLET 81
Qy 74 GPVLAQMGHGEHSAKRRIIVRSFTIGDALDHLSPKQENLAPYLERGKSLDVLNDFGK 133
Db 82 RIMNSTDPHEHNRVSRIVSKATPFRSLEQWKPRIOAIANELVQHICKYSEVNIVEFAA 141
Qy 134 TFACVVTMDLGLDKRDHEKISEHGWGVADPITSQSPEARAHSLWCSEQLSOYLMPIV 193
Db 142 PLPVTWISDLGLVPTTDRKKIKAW-SDILEWPMYSKEKFNLDLVEKGIALNEFKAYLLPIV 200
Qy 194 KERVNPGSLITSLCTSEYEGWALSQDILALINVLNLAATPDKTALMIY-HLNN 252
Db 201 QEKRYHLTDDIISLIRAEYEGERLTDEEIVTSLGLAAGNETTNLIINSFYCFLVDS 260
Qy 253 PEQWMDVLRSLVPRAIAETLRYKPPVQLIPRLSQDTVVGGMEIKKDTIVFCMIGAA 312
Db 261 PGTYKELREPTLISKAIEVLVRFPIITL-ARRITEDTNIFGLMKKQDQMVVAVWSAAN 319
Qy 313 RDEAFQPDVFNHREDLGIKAFSGAARHLAPFGSGIHNCVGAAPAKNEIEIVANIVLD 372
Db 320 LDEKKFSQSKFNHR-----IGNEKHLTFGKGPHFCGLAPLARLEAETALTFFIN 370
Qy 373 KMRNIRLEEDF----CYAES 388
Db 371 AFEKIALSPSFNLEQCILEN 390
RESULT 15
Q83X67
ID Q83X67 PRELIMINARY; PRT; 404 AA.
AC Q83X67;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE P450-like hydroxylase
OS Streptomyces rochei (Streptomyces parvullus).
OG Plasmid pSLA2-L.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1928;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=7434AN4;
RX MEDLINE=22676866; PubMed=12791134;
RA Mochizuki S., Hiratsu K., Suwa M., Ishii T., Sugino F., Yamada K.,
RA Kinashi H.;
RT "The large linear plasmid pSLA2-L of Streptomyces rochei has an
RT unusually condensed gene organization for secondary metabolism.";
RL Mol. Microbiol. 48:1501-1510(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=7434AN4;
RX MEDLINE=20408175; PubMed=10954087;
RA Hiratsu K., Mochizuki S., Kinashi H.;
RT "Cloning and analysis of the replication origin and the telomeres of
RT the large linear plasmid pSLA2-L in Streptomyces rochei.";
RL Mol. Gen. Genet. 263:1015-1021(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=7434AN4;
RX MEDLINE=20321737; PubMed=10767533;
RA Suwa M., Sugino H., Sasaoka A., Mori E., Fujii S., Shinkawa H.,
RA Nimi O., Kinashi H.;
RT "Identification of two polyketide synthase gene clusters on the linear
RT plasmid pSLA2-L in Streptomyces rochei.";
RL Gene 246:123-131(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA STRAIN=7434AN4;
RX MEDLINE=99053144; PubMed=9836424;
RA Kinashi H., Fujii S., Hatanai A., Kurokawa T., Shinkawa H.;
RT "Physical mapping of the linear plasmid pSLA2-L and localization of
RT the eryal and acti homologs.";
RL Biosci. Biotechnol. Biochem. 62:1892-1897(1998).
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
DR ENBL; AB088224; BAC76495.1; -.
DR HSSP; Q00441; 10XA.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002397; BP450.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00359; BP450.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN 1.
DR Heme; Monooxygenase; Oxidoreductase; Plasmid.
SQ SEQUENCE 404 AA; 45340 MW; 1C9C00F7E7909891 CRC64;
Query Match 20.8%; Score 433.5; DB 2; Length 404;
Best Local Similarity 26.8%; Pred. No. 6.2e-23;
Matches 113; Conservative 77; Mismatches 169; Indels 63; Gaps 9;
Qy 12 SDQFQNNPAYFSQLEEDPV-----HYESIDSYFISRVHVDVYILQHP-----D 57
Db 14 SELFHTDQATVRYEILEQRFPVTRVFDGSL-VMLVNRHEDVRAALTDPRLSNDPMKQSD 72
Qy 58 I-----FTTKSLVERAEPVVRGPGVLAQMGHGEHSAKRRIIVRSFTIGDALDH 103
Db 73 IDLSAATGIPADLIEYFORNMFSDP-----DHGRLRLKLVTRFETVRINA 119
Qy 104 LSPLIKQENLAPYLERGKSLDVLNDFGKTFAVCVTMDMLGDKRDHEKISEHWSGVAD 163
Db 120 LRPRIRQIADDDLEKFAANTGGDLVEALARPLPTVMCELLGVPEEDRADFTWSQHI-- 177
Qy 164 FITSISQSPPEARAHSLWCSEQLSOYLMPIVKERRVNPVSGSLISILCTSEYEGWALSQDI 223
Db 178 ----VESSPEFARNVAVSVRSLEFECVRSILRRRDEPGDLSALVLDRLDVAORLSNEL 233
Qy 224 LALILNVLAAATEPADKTLALMIYHLNINPEQWMDVLRSLVPRAIAETLRYKPPVOLI 283
Db 234 ISTVFLVLLAGIETTNNVLTGTGTFLLLTHPGELARLADGALLGPVAEEMLRVAPIET 293
```

```

QY 284 PROLSQDTVVGGMEIKDITIVFCMIGAANRDPFAFQPDVFNHREDLGKSAFSGAARH 343
Db 294 SRHTEPVEIGVSDAQSTVLINLAAANRDPARFEDPQSFVRVDRNDGG-----H 343
QY 344 LAFGSGIHNCVGAFAKNEIEIVANIVLDKMRNRL---EEDFCYAESGLYTGVPVSLLY 400
Db 344 LTFGHGHIYCLGAALARAEEVTFEALLERFDLRLAASASDLTWHA--FMRGPVELPV 401
QY 401 AF 402
Db 402 SW 403

```

Search completed: December 10, 2004, 13:44:56
 Job time : 159.579 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2004, 13:25:54 ; Search time 30.2515 Seconds
(without alignments)
1288.126 Million cell updates/sec

Title: US-10-627-124-2
Perfect score: 2080
Sequence: 1 MSQSIXLFSVLSDQFQNNPY.....ABSGLYTRGPVSLLVAFDGA 405
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues 283416
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	405	1 F69611	cytochrome P450 cy
2	468	22.5	405	1 B42606	cytochrome P450 cv
3	424.5	20.4	400	2 E82817	cytochrome P-450 h
4	418.5	20.1	396	1 B69851	cytochrome P450 xj
5	415.5	20.0	397	1 S51594	cytochrome P450 mj
6	409.5	19.7	410	2 S39924	cytochrome P450meg
7	408.5	19.6	402	1 A70707	cytochrome P450 rv
8	407	19.6	417	1 S49051	cytochrome P450 ty
9	400.5	19.3	410	1 B69611	cytochrome P450 cy
10	399.5	19.2	402	2 H82813	cytochrome P450-11
11	399.5	19.2	411	2 T36529	probable cytochrom
12	388.5	18.7	376	1 G69679	polyketide hydroxy
13	387	18.6	410	1 O4856M	cytochrome P450 bi
14	383.5	18.4	395	1 G69594	cytochrome P450 bo
15	367.5	17.7	411	1 JC4003	cytochrome P450 -
16	359	17.3	412	1 B40634	erythromycin monoo
17	348.5	16.8	434	2 C87170	probable cytochrom
18	347.5	16.7	406	2 A35401	cytochrome P450 10
19	346.5	16.7	415	2 B97512	probable cytochrom
20	346.5	16.7	415	2 AC2731	cytochrome P450 by
21	345.5	16.6	406	1 S18531	cytochrome P450 er
22	343.5	16.5	428	1 F70729	cytochrome P450 rv
23	343	16.5	411	1 JCS989	polyketide synthas
24	338	16.2	398	1 H70807	cytochrome P450 rv
25	335	16.1	403	2 B35401	cytochrome P450 li
26	333.5	16.0	310	2 T44767	cytochrome P450 1i
27	329	15.8	438	1 E70515	cytochrome P450 rv
28	328	15.8	407	2 B32306	cytochrome P450 cy
29	326.5	15.7	381	1 S15809	cytochrome P450 CY

make gene
day
teach
+
mily

RESULT 1
F69611
Cytochrome P450 CYPX - Bacillus subtilis
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: F69611
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter...

ALIGNMENTS

30	325	15.6	402	2	JC5151	nitric-oxide reduc
31	325	15.6	404	1	JC5150	nitric-oxide reduc
32	320	15.4	434	1	B70511	cytochrome P450 rv
33	319.5	15.4	489	1	H70729	cytochrome P450 -
34	318.5	15.3	404	2	T30231	cytochrome P450 fa
35	316	15.2	444	2	F87366	cytochrome P450 rv
36	315	15.1	433	1	B70677	cytochrome P450 fa
37	313	15.0	424	2	E87558	cytochrome P450 sc
38	312	15.0	410	2	JC4287	linalool 8-monooxy
39	311	15.0	406	1	A48495	cytochrome P450ter
40	311	15.0	428	1	A42971	cytochrome P450 no
41	310.5	14.9	408	2	PD0007	cytochrome P450 no
42	310.5	14.9	408	2	JC5674	cytochrome P450 no
43	305.5	14.7	399	1	A55578	cytochrome P450 -
44	305	14.7	467	2	T10879	y41C protein - Rhi
45	301	14.5	391	2	A63246	P-450 monooxygenase

Query Match 100.0%; Score 2080; DB 1; Length 405;
Best Local Similarity 100.0%; Pred. No. 2.1e-149;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSQSIXLFSVLSDQFQNNPYFSQLREDPVHYEESIDSYFISRVHYVRILOHPDIFT 60
DB 1 MSQSIXLFSVLSDQFQNNPYFSQLREDPVHYEESIDSYFISRVHYVRILOHPDIFT 60
QY 61 TKSLEVERAEPVWRGVLQAQHGKHSKAKRIIVRSFIGDALHSLIKONAEHLAPYL 120
DB 61 TKSLEVERAEPVWRGVLQAQHGKHSKAKRIIVRSFIGDALHSLIKONAEHLAPYL 120
QY 121 ERGKSLDNLVDFGKTFAVCVTMDMLGDKRDHEKISEHWSGVADFTISISQSPARASLW 180
DB 121 ERGKSLDNLVDFGKTFAVCVTMDMLGDKRDHEKISEHWSGVADFTISISQSPARASLW 180
QY 181 CSEQLSQYLMPIVKERRVNPVPGSDLSILCTSEYEGMALSDKOILAILNLVLAATSPADK 240

Db 181 CSEQLSQYLMPIKERRVNPGLDLSILCTSEYEGWALSDKDLILALLNVLAAATEPADX 240
QY 241 TLALMIYHLNPNQNDVLADSLVPRAIAETRLRYKPPVQLIPQLSDTIVVGWMEIKK 300
Db 241 TLALMIYHLNPNQNDVLADSLVPRAIAETRLRYKPPVQLIPQLSDTIVVGWMEIKK 300
QY 301 DTVFCMIGAAANDPFAFPQDVFNHREDLIGKSAFSGAARHLAFSGGIHNCVGAAPAK 360
Db 301 DTVFCMIGAAANDPFAFPQDVFNHREDLIGKSAFSGAARHLAFSGGIHNCVGAAPAK 360
QY 361 NEIEIVANIVLDMKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405
Db 361 NEIEIVANIVLDMKRNIRLEEDFCYAESGLYTRGPVSLVAFDGA 405
RESULT 2
B42606
cytochrome P450 CYP1B1 - Saccharopolyspora erythraea
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Saccharopolyspora erythraea
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B42606
R:Andersen, J.F.; Hutchingson, C.R.
J. Bacteriol. 174, 725-735, 1992
A:Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and enzy
A:Reference number: A42606; PMID:19211109; PMID:1732208
A:Accession: B42606
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA; protein
A:Residues: 1-405 <AND>
A:Cross-references: UNIPROT:P33271; GB:M83110; NID:gl52682; PIDN:AAA26483.1; PID:gl52684
A:Experimental source: NRR12338
A:Note: sequence extracted from NCBI backbone (NCBIP:77484)
C:Genetics:
A:Gene: CYP107B1
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:238-374/Domain: cytochrome P450 homology <CYP>
F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted
Query Match 22.5%; Score 468; DB 1; Length 405;
Best Local Similarity 30.0%; Pred. No. 1.1e-27;
Matches 115; Conservative 85; Mismatches 149; Indels 34; Gaps 8;
QY 13 DQFQNNPYAYFSQLRPE--DPVHYEESIDSYFISRVHVRVYLQHP-----DIFTTKSLVE 66
Db 14 DAFAQDRHNRYARMREBPQVIRTVNGLDAWLITRYEDVKALLDPRIAKDFGRTQOITE 73
QY 67 R--AEPVMR-----GPVLAQMKGKHSKRRIVRSFIGNALDHLSPLIKQNAENLLA 117
Db 74 KKLADARPGFSPLDGLPHMLNTDPPDHLRLKLVVKAFTARRVEGLRPRIQITDLDLD 133
QY 118 PYLERKSLDVLNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVADFTISQSPEARAH 177
Db 134 RLAGSEVDLIDFAPPLPITWISLGMVEDSRDRDFTSRWTNVLD-----GSQPEAQAO 188
QY 178 SLWCSQSLSQYLMPIKERRVNPGLDLSILCTSEYEGWALSDKDLILALLNVLAAATEP 237
Db 189 A---SVAMVEYLTELLAKRTPEGDDLTALEAVEGDGRLSEGELIAMVFLLVAGHET 245
QY 238 ADKTLALMIYHLNPNQNDVLADSLVPRAIAETRLRYKPPV-QLIPQLSDTIVVGW 296
Db 246 TVNLGNCSVLISGNPDQAAARNDPSLLPFAETLRYESPVSVAAGTFRHTAEAVRFGDV 305
QY 297 EIKKDTIVFCMIGAAANDPFAFPQDVFNHREDLIGKSAFSGAARHLAFSGGIHNCVGA 356
Db 306 VIPEGLVWVALGAANDGDERFDDPRFDITRETTG-----HVAFGHGIHFCVGA 355
QY 357 AFANKEIEIVANIVLDMKRNIRL 379
Db 356 ALARLEAQIAVGRLLRRFPDLRM 378

RESULT 3
E82817
cytochrome P-450 hydroxylase XF0356 [imported] - Xylella fastidiosa (strain 9aSc)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: E82817
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82817
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-400 <SIM>
A:Cross-references: UNIPROT:Q9PGE6; GB:AB003887; GB:AB003849; NID:g9105175; PIDN:AAF8316
A:Experimental source: strain 9aSc
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
Briões, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Carraro, D.M.; Carver, H.
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.B.; Laigre
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Marino, C.L.; Marques, M.V.; Martins, E.
A:Authors: Martins, E.M.P.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Teuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0356
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
F:235-370/Domain: cytochrome P450 homology <P45>
Query Match 20.4%; Score 424.5; DB 2; Length 400;
Best Local Similarity 28.0%; Pred. No. 2.1e-24;
Matches 111; Conservative 78; Mismatches 164; Indels 43; Gaps 6;
QY 5 IKLFSVLSDQFQNNPYAYFSQLRPEDPVHYEESIDSYFIS-----RYHDVRYILOH 55
Db 1 MKLADLSSPAFLNYPPLYETLRQ-----GSFVSIQGNALMTGRIYSIVDGLLHN 50
QY 56 PDIFTTKSLVERAE-----PVMRG--PVLQMGKHSKRRIVRSFIGNALDHL 103
Db 51 RNM--GKSYMESIRVRYGDDALDMLPFCQFNRMFLMLNPPVHTLRLGLVMQAFTRGSES 108
QY 104 LSPLIKQNAENLLAPYLERKSLDVLNDFGKTFACVCTMDMLGDKRDHEKISEWHSGVAD 163
Db 109 MKPLATDTAHLRDDFEQKSSVDLTFEFPPLPMRIICMMDDVDISDAISLVSVAENIAK 168
QY 164 FITTSQSPEARAHSLWCSEQSLYLMPIKERRVNPGLDLSILCTSEYEGWALSDKDI 223
Db 169 VLDPAWSPDELVHASAAVEELAHYFTRLIELRRAPQGTDLISMLLRAEEGQKLTHDEI 228
QY 224 LALIINVLAAATEPADKTLALMIYHLNPNQNDVLADSLVPRAIAETRLRYKPPVQLI 283
Db 229 VSNVTLLIGGYETTSNMIGNALIALHRHPKQARLKSDDLMLPQAILCECLRYDGSVQFT 288
QY 284 PQLSQDITVVGWMEIKKDTIVFCMIGAAANDPFAFPQDVFNHREDLIGKSAFSGAARH 343
Db 289 MRAAMDDVSIQGVVFRGTIVFLMIGAAANDPFAQTDDPHLDITRKQ-----GRL 338
QY 344 LAFSGGIHNCVGAAPAKNEIEIVANIVLDMKRNIRL 379
Db 339 QSGFGAVHHCIGYRLALVELECATVLLERLPHRL 374
RESULT 4
B69851
cytochrome P450 yjib - Bacillus subtilis

N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus subtilis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B69851
R:Kunzt, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux, C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choudhury, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallenbach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Mausel, Y., M.; Osawa, K.; Ogiwara, A.; Oudeg, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Schlegel, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Tanaka, T.; Terpe, T.; Terpe, A.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, K.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; PMID:98044033; PMID:9384377
A:Accession: B69851
A:Status: nucleic acid sequence not shown; translation not shown
A:Keywords: oxidoreductase
F:236-371/Domain: cytochrome P450 homology <CYP>
Query Match 20.1%; Score 418.5; DB 1; Length 396;
Best Local Similarity 26.2%; Pred. No. 5.8e-24;
Matches 95; Conservative 86; Mismatches 155; Indels 27; Gaps 6;
QY 18 NPAYFSQREEDPVHYESIDSYSFISRDVRYIYLOHPDFTTKSLVERAPVWGPNVL 77
DB 25 HPPFPVYEMRKAPVDFEENQVMSVFLYDDVKVGVGDLFSS-CMPQQTSSI--GNSI 81
QY 78 AQMHGKHSKRIIVRSFIGDLDHLSPLIKONAENLLAPYLERGK-SDLVNDFGK 137
DB 82 INMDPPKTKIRSVNKAFTPRVKNQWEPRIQITDELQKFGSRSEFDLVHDFSPVLPV 141
QY 138 CVTMDMLGLDKRDHEKISEHSGVADFTISQSPPEARAHSLW-----CSEQLSYLMP 191
DB 142 IVISELLGVPSAHMEQFKW----SDLLVSTPKDSEAEKAFLEERDKCEELAAFPAG 197
QY 192 VIKERRVNPFGSLLISILCTSEYEGMALSKDILAILNVLAAATEPADKTLALMIYHLN 251
DB 198 IIEKRNKPEQDIISILVEAEETGKLSGELIPFCTLLVAGNETTTNLISNAMYSL 257
QY 252 NPEQMDVLADSLVPRATLRYPPVQLIPRQLSODTVVGGMEIKKDTIVFCMIGAA 311
DB 258 TPGVVEELSHPELMPQVVEALRFAPAPVLRRIAKRDTGIGHLIKGDMVLAFAVA 317
QY 312 NRDPFAEPQDFVNIHREDLIGKISAFSGAARHLAFSGHNCVGAFAFAKNEIHEIVANVL 371
DB 318 NRDEAKFDRPHMFDIRH-----PNPHAFGHGHFCLGAPLARLE-----ANIAL 363
QY 372 DKM 374
DB 364 TSL 366
RESULT 5
S51594
N:Contains: Micromonospora griseorubida
C:Species: Micromonospora griseorubida
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S51594
R:Inouye, M.; Takada, Y.; Muto, N.; Beppu, T.; Horinouchi, S.

Mol. Gen. Genet. 245, 456-464, 1994
A:Title: Characterization and expression of a P-450-like mycinamicin biosynthesis gene u
A:Reference number: S51593; PMID:95107242; PMID:7808395
A:Accession: S51594
A:Molecule type: DNA
A:Residues: 1-397 <INO>
A:Cross-references: UNIPROT:Q59523; EMBL:D16098; NID:g286050; PIDN:BA03672.1; PID:g3036
C:Genetics:
A:Gene: mycG
A:Start codon: GTG
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: heme; oxidoreductase
F:231-368/Domain: cytochrome P450 homology <CYP>
Query Match 20.0%; Score 415.5; DB 1; Length 397;
Best Local Similarity 28.9%; Pred. No. 9.8e-24;
Matches 114; Conservative 76; Mismatches 171; Indels 33; Gaps 11;
QY 23 FSQLEEDPVH-----YESIDSYSFISRDVRYIYLOHPDFTTKSLVE---RAEPVMRG 74
DB 23 YGELQETEPVSRPPYGE--EAWLVTRYEDVRAVLGDRGFRVGRPSMTRDEPRTRPEMVK 80
QY 75 PVLAQMHGKHSKRIIVRSFIGDLDHLSPLIKONAENLLAPYLERGK-SDLVNDFGK 133
DB 81 GGLSMDPPEHSLRLRLVVKATARRAESLRPRAREIAHELVDQMAATQPADLVAMFAR 140
QY 134 TFAVCVTMDMLGLDKRDHEKISEHSGVADFTISQSPPEARAHSLWSEQLSYLMPVI 193
DB 141 QLPVRVICELGVPASADHDFRTRW-SGA--FLSTAETVAEMQEA--AEQAYAYMGDLI 194
QY 194 KERRVNPFGSLLISILCTSEYEGMALSKDILAILNVLAAATEPADKTLALMIYHLN 253
DB 195 DRRRKEPTDVLVSALVQARDQDLSLSEQLLGLVAGVYESTTTQIADFYVLLMTRP 254
QY 254 EQMDNDVLADSLVPRATLRYKPPVQLIPRQLSODTVVGGMEIKKDTIVFCMIGAA 311
DB 255 ELRRQLLDREPLIPSAVEELTRVPLGVTGTAFFRAVEDVTLRGVTIRAGEPVLAAGAA 314
QY 312 NRDPFAEPQDFVNIHREDLIGKISAFSGAARHLAFSGHNCVGAFAFAKNEIHEIVANVL 371
DB 315 NRDAQFPDADRIDVDRT-----PNQHLGFGHGHVHCLGAPLARVQLVALEVL 364
QY 372 DKMIRNL---BEDFCYAESGLYTRGPVSLVAF 402
DB 365 QRLPGIRLIGPITQLRWSE-GMLLRGPLELPVW 397
RESULT 6
S39924
Cytochrome P450meg - Bacillus megaterium (ATCC 13368)
C:Species: Bacillus megaterium
A:Variety: ATCC 13368
C>Date: 07-Oct-1994 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C:Accession: S39924; S32216
R:Kraehenbach, R.; Isernhagen, M.; Noeske-Jungblut, C.; Boidol, W.; Siewert, G.
Mol. Gen. Genet. 241, 170-176, 1993
A:Title: Cloning sequencing and expression of the gene for cytochrome P450meg, the stereo
A:Reference number: S39924; PMID:94049677; PMID:8232201
A:Accession: S39924
A:Molecule type: DNA
A:Residues: 1-410 <RAU>
A:Cross-references: UNIPROT:Q06069; EMBL:221972; NID:g288298; PIDN:CAA79985.1; PID:g2883
A:Experimental source: ATCC 13368
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:240-377/Domain: cytochrome P450 homology <P45>
F:355/Binding site: heme iron (Cys) (axial ligand) #status predicted
Query Match 19.7%; Score 409.5; DB 2; Length 410;
Best Local Similarity 26.8%; Pred. No. 2.9e-23;
Matches 109; Conservative 85; Mismatches 176; Indels 37; Gaps 9;
QY 3 QSIKFLSVLSQDFQNNFYAFYFQLEEDPVHYESIDSYSFISRDVRYIYLOHPDIF--- 59

RESULT 9

E69611
C;Accession: E69611; T44774
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Species: *Bacillus subtilis*
N;Contains: oxidoreductase (EC 1.-.-.-)
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton
C.; Bron, S.; Brouillette, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallet
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.P.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, K
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9394377
A;Accession: E69611
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-410 <KUN>
A;Cross-references: UNIPROT:O08469; GB:Z99117; GB:AL009126; NID:g2634966; PIDN:CAB14615.
R;Bellcsy, B.R.; Gustafsson, M.C.U.; Sonenshein, A.L.; von Wachenfeldt, C.
J. Bacteriol. 179, 5448-5457, 1997
A;Title: An lrp-like gene of *Bacillus subtilis* involved in branched-chain amino acid tra
A;Reference number: Z22837; MUID:97431495; PMID:9287000
A;Accession: T44774
A;Status: translated from GB/ENBL/DBD
A;Molecule type: DNA
A;Residues: 1-410 <BEL>
A;Cross-references: EMBL:Y11043; NID:g1926275; PIDN:CAA71937.1; PID:g1926278
A;Experimental source: strain 1A1
C;Genetics:
A;Gene: *cypA*
C;Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology
C;Keywords: oxidoreductase
F;245-381/Domain: cytochrome P450 homology <CYP>

Query Match 19.3%; Score 400.5; DB 1; Length 410;
Best Local Similarity 26.2%; Pred. No. 1.4e-22;
Matches 105; Conservative 95; Mismatches 162; Indels 39; Gaps 9;

QY 2 SQSIKFLSVLSQ-----FQNNPYAFSOLREEDPVHYBESIDS-----YFISRYHDVRY 51
DB 3 SKEKSVTILTESQSSRAFDEAYEFYKELKRSQAL-YPLSLGALCKGWLISRYDDAIH 61
QY 52 ILQ-----HPDIPTTK-----SILVERAEPVMPGVLVAQMHGKHSKRIVRSPTGDA 100
DB 62 LLKNEKLKNVENVFTAKEKRPALLKNEETLTKH--MLNSDPDPHNRRLTLVQKAFTHRM 119
QY 101 LDHLSPLIKQNAENLLAPYLERGKSDLVNDFGKTFACVCTMDMLGLDKRHEKISEWHSG 160
DB 120 ILQLEDKIQHADSLLDKVQNKPMNVLDVYAFPLPIIVISEMLGIPLEDQRQKPRVWSQA 179
QY 161 VADPITTSISQSPARASHLWCSEQLSOYLMPVIERKRVNPGSDLSILCTSEYEGMALSD 220
DB 180 IIDF-----SDAPERLOENDHLLGFEVYLSLVKRRKREPAGDLSALIQAESEGTQLST 235
QY 221 KQIALILNVLAAEPADKTLALMIYHLLNPNQMNVDLADSLVPRAIAETLRYKPPV 280
DB 236 EELYSMIMLLIVAGHETTVNLITNMTYALMCHHDQLEKLRQQDPDMNSAIEEALRPHSPV 295
QY 281 QLP-RQLSODTVVGGMEIKKOTIVFCMIGANRDPPEAFQPDVFNTHREDLGKISAFSG 339
DB 296 ELTIRWTAEFFILHGOEIKRKVDIIITSLASANRDEKIFPNADIFDERK----- 346

QY 340 AARHLAFSGIHNVCVGAAPAKNAEIEIVANVLDKMRNIRLE 380
DB -NRHIAFGHGNHFCGLGAQLARLEAKIAISTLLRCPCNIQLK 386

RESULT 10

H82813
C;Accession: H82813
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Species: *Xylella fastidiosa*
N;Contains: cytochrome P450-like enzyme XP0377 [imported] - *Xylella fastidiosa* (strain 9a5c)
R;Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: H82813
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-402 <SIM>
A;Cross-references: UNIPROT:Q9PGC5; GB:AE003889; GB:AE003849; NID:g9105205; PIDN:AAF0318
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohne
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XP0377
C;Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology
F;235-370/Domain: cytochrome P450 homology <P45>

Query Match 19.2%; Score 399.5; DB 2; Length 402;
Best Local Similarity 27.7%; Pred. No. 1.6e-22;
Matches 110; Conservative 76; Mismatches 166; Indels 45; Gaps 6;

QY 5 IKFLSVLSQDFQNNPYAFSOLREEDPVHYEESIDSYFIS-----RYHDVRYILOH 55
DB 1 MKLTDLSNPAFLNPPLVETLRAQP-----FVSGPNALMTGYSYLVDSLLHN 50
QY 56 PDFTTKSLVERAEPVNRGVPVLAQM-----HKGKHSKRIRIVRSGIFGDALD 102
DB 51 RNM--GKYMESMR-VRYGDSAADMPLFOAFSRMFTINPPAHTLRLGLVQAFTGRESE 107
QY 103 HLSPLIKQNAENLLAPYLERGKSDLVNDFGKTFACVCTMDMLGLDKRHEKISEWHSGVA 162
DB 108 SMRPLAIDTAHQIDNFEQKPSVDLVAEFAPFPFPMQIQICMDMDVDIGDAVTLGTAFAKIA 167
QY 163 DFITSISQSPARASHLWCSEQLSOYLMPVIERKRVNPGSDLSILCTSEYEGMALSDKD 222
DB 168 KVFDPSPMSADELVHASTAYEEELAQYFTKUIERRTHPGTDLISMFRAEDGEKLTDE 227
QY 223 ILALINVLAAEPADKTLALMIYHLLNPNQMNVDLADSLVPRAIAETLRYKPPVOL 282
DB 228 IVSNVIMLLIAGYETTSSNMIGNALIAHRRHQELALLKSDLSLMPQAVSCLRYDGSVQF 287
QY 283 IPQLSODTVVGGMEIKKOTIVFCMIGANRDPPEAFQPDVFNTHREDLGKISAFSGAAR 342
DB 288 TMAAMDIDIEVEGELVPRGTWVFLMGLGANRDPQAQFTHPDQLDITRQK-----GR 337
QY 343 HLAFSGIHNVCVGAAPAKNAEIEIVANVLDKMRNIRL 379
DB 338 LQSGFAGIHHCGLGYRLALIELECATLTLFLERLPHRLR 374

RESULT 11

T3526

probable cytochrome P450 hydroxylase - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C/Accession: T36526

R/Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1999

A/Reference number: Z21609

A/Accession: T36526

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-411 <MUR>

A/Cross-references: UNIPROT:Q9X8Q3; EMBL:AL049754; PIDN:CAB42023.1; GSPDB:GN00070; SCORE

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: SCODB:SCH10.14C

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C/Keywords: heme; iron; metalloprotein

P;250-385/Domain: cytochrome P450 homology <P45>

F;363/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 19.2%; Score 399.5; DB 2; Length 411;

Best Local Similarity 27.7%; Pred. No. 1.7e-22;

Matches 106; Conservative 75; Mismatches 169; Indels 33; Gaps 7;

QY 15 FQNNPYAYFQSLREDDPVHYEISDSYFISRYHDVRYILO-----HPDIFTT 61

DB 22 FVADPYPAFAELRARGRLYFEPDQWLVPHHADVSALLDRDRIGRTYQHRFTHEDFGT 81

QY 62 KSLVERAEP--VMRGPVLAQMKGKHSKRIIVRSPIGDALDHLSPKLIQKNAENLLAPY 119

DB 82 PPPPEQ-EPHTLNDHGMLDLEPPDHTRIIRLVSKAFTPTVERKPYVHGLADDLVARL 140

QY 120 LERKSLVNDVFGKTFVAVCTMDMLGDKRDHEKISEHSGVADFTITSQSPEARHSL 179

DB 141 VAAGGDLTDVAEPLPVAVIAEMLGIPESDRAPLRPWSAICG-MYELNPSEETAAKAV 199

QY 180 WCEQLSOYLMPIKERRVNPVSGDLISILCTSEYEGNALSDKOLA---LILNVLLAATE 236

DB 200 RASLDFSDYLRALFAARKEPFGDDLISGLIAAHDEDDLRTQEIMISTCVLLLNAGHEATV 259

QY 237 PADKTLALMTVYHLNPNQMDVLADSLVPRAETLRYKPPVQLIPROLSDQTVVGGM 296

DB 260 NATNGWLAF--RHFDQLAALRADHSLVPSAVEELMRDTPQLFERWVLEIDEIDGT 316

QY 297 EIKKDTIVFCMIGANRDPPEAFQPDVFNTHREDLGKSAFSGAARHAFSGSHNCVGA 356

DB 317 TLPRAEAVMLFGSANHPAVFTDPERLDLTRD-----NPHISFSAGIHYCIGA 366

QY 357 AFAPKNEIEIVANVLDKMRNRL 379

DB 367 PLARIELAASMTSLKRAPGLRL 389

RESULT 12

G69679

polyketide hydroxylase (EC 1.-.-.-) pks - Bacillus subtilis

C/Species: Bacillus subtilis

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C/Accession: G69679

R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

Koester, P.; Koningsstein, G.; Kroch, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois

A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Nasuda, S.; Mauel

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yamoto, K.; Yata, K.; Yoshida, K. A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A/Reference number: A69580; MUID:98044033; PMID:9384377

A/Accession: G69679

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-376 <KUN>

A/Cross-references: UNIPROT:O31785; GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13607.1

A/Experimental source: strain 168

C/Genetics:

A/Gene: pksS

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C/Keywords: oxidoreductase

F;240-376/Domain: cytochrome P450 homology <CYP>

Query Match 18.7%; Score 388.5; DB 1; Length 376;

Best Local Similarity 24.6%; Pred. No. 1e-21;

Matches 98; Conservative 80; Mismatches 161; Indels 59; Gaps 7;

QY 7 LFSVLSDQFQNNPYAYFSQLREDDPVH-----FTTKSLVERAEPVMRGPVLAQMKGKHSKRIIVRSFI 97

DB 7 MFPHGKGFHHNPFVSLGRFREEPFHRFELKRFGATYPAWLITRYDDCMFAFLKDNRI 66

QY 48 DVRYIILQHPDI-----FTTKSLVERAEPVMRGPVLAQMKGKHSKRIIVRSFI 97

DB 67 DVKNVNNQIKMLNVSEIDIFVSDHMLAKDTP-----DHTFLRSILVHQAFT 113

QY 98 GDALDHLSPKLIQKNAENLAPYLERGKSLVNDVFGKTFVCTMDMLGDKRDHEKISEW 157

DB 114 PRTIENLRGSIQIEAQLDEMEKENKADIMKSFASPLPFIVISLWMLGIPKEDRSQFOIW 173

QY 158 HSGVADFTISISQSPARASHLWCSEQLSOYLMPIKERRVNPVSGDLISILCTSEYEGMA 217

DB 174 TNAWVD-----TSEGNRELTNOALREFKDYIAKLHIDRIKPKDDLISKLVAENGSK 227

QY 218 LSDKDLALILNVLLAATPADKTLALMTYHLLNPNPEQMDVLADSLVPRATAETLRYK 277

DB 228 LSEKELYSMFLVAVAGLETTVNLGSGTLLALQHKKECKLQKQPEMIATAVEELLRYT 287

QY 278 PPVOLIP-RQLSDQTVVGGMEIKKOTIVFCMIGANRDPPEAFQPDVFNTHREDLGKSA 336

DB 288 SPVVMANRWAIEDFTYKSHIKRGDMIFIGISANRDPNFFENPEILNIR----- 339

QY 337 FSGAAHRLAFSGSHNCVGAAPAKNIEIVANVLDK 374

DB 340 --SPNRHISFGFGIHFLGAPLARLEGHTAFKAFFEEI 375

RESULT 13

O4BS6M

cytochrome P450 106 - Bacillus megaterium

N/Alternate names: cytochrome P450BM-1

N/Contains: oxidoreductase (EC 1.-.-.-)

C/Species: Bacillus megaterium

C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004

C/Accession: S07764; S17973

R/He, J.S.; Ruettinger, R.T.; Liu, H.M.; Fulco, A.J.

Biochim. Biophys. Acta 1009, 301-303, 1989

A/Title: Molecular cloning, coding nucleotides and the deduced amino acid sequence of P-

A/Reference number: S07764; MUID:90089408; PMID:2597681

A/Accession: S07764

A/Molecule type: DNA

A/Residues: 1-410 <HEJ1>

A/Cross-references: UNIPROT:P14762; EMBL:X16610; NID:g39626; PIDN:CAA34612.1; PID:g39627

A/Accession: S17973

A/Molecule type: protein

A/Residues: 1-25 <HEJ2>

C/Genetics:

A/Gene: CYP106

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;

2000

Db	233	EETATLVTFTAGNETTAHLIGNGVAALITHSDQLALLRSRRLLSQAVDELMRWCTPV	292
Qy	281	OLI-PROLSODTVVGGMEIKKDTIVFCMIGAAARDPEAFQDPDVFNHREDLGKSAFSG	339
Db	293	QVTQPRYATEDLDVGGVTVRKGEQVAVIGAAGHDPRFPDPERFDITRNHRAPHEA---	349
Qy	340	AARHLAFSGGIHNCVGAFAKNEIEIVANIVLDKMRNRL	379
Db	350	---HVGFGFGPHYCLGAALAHQETAIALDTLDFRFPFLAL	386

Search completed: December 10, 2004, 13:45:49
Job time : 31.2515 secs